

Biologieunterricht im Naturkundemuseum
im Spannungsfeld zwischen Instruktion und Konstruktion
– eine empirische Untersuchung zu kognitiven und affektiven Lerneffekten
(am Beispiel des Umweltschutz-Informationszentrums Lindenhof in Bayreuth)

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von
Matthias Wilde
aus Bayreuth

Juni 2004

Die vorliegende Arbeit wurde vom Oktober 2001 bis Juni 2004 am Lehrstuhl Didaktik der Biologie der Universität Bayreuth angefertigt.

Vollständiger Abdruck der von der Fakultät Biologie, Chemie und Geowissenschaften genehmigten Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaft (Dr. rer. nat.).

Tag der Einreichung: 07.06.2004

Tag des wissenschaftlichen Kolloquiums: 03.08.2004

1. Gutachter: Prof. Dr. S. Klautke

2. Gutachter: Prof. Dr. K. Dettner

Prüfungsausschuss: Prof. Dr. E. Steudle
Prof. Dr. F.X. Bogner
Prof. Dr. K. Fiedler

“Tell me, boy, does anything penetrate that thick skull of yours? Didn’t you hear me say, quite clearly, that only one rat spleen was needed? Didn’t I state plainly that a dash of leech juice would suffice? What do I have to do to make you understand, Longbottom?” –

Prof. Snape zu dem Zauberschüler Neville Longbottom, ein Zitat aus “Harry Potter and the Prisoner of Azkaban“ von J. K. Rowling (1999).

Selbst in der Phantasie einer fiktiven Zauberschule herrscht instruktionsbestimmter Unterricht vor so wie in realen Schulen Deutschlands, den meisten Ländern Europas und Nordamerikas (Gruber et al. 2000, S. 140). Trotzdem stellt man auch hier (aus den falschen Gründen) die richtige Frage, eine der zentralen Fragen der Didaktik:

Was kann man (als Lehrer) unternehmen, damit Schüler verstehen (vgl. Ladenthin 2001, S. 307)?

I.	EINLEITUNG	1
1.	Zum Forschungsgegenstand	1
2.	Lerntheoretische Grundlagen: Konstruktion und Instruktion	5
2.1.	Aktueller Theoriekontext der Lerntheorien	6
2.1.1.	Konstruktivismus	6
2.1.2.	„Instruktivismus“	11
2.1.3.	Fazit	12
2.2.	Historischer Kontext von Instruktion und Konstruktion	12
2.2.1.	Zur Tradition der Instruktion	13
2.2.2.	Zu Vorläufern konstruktivistischer Positionen	15
2.2.3.	Fazit	17
2.3.	Aktuelle Lerntheorien des Konstruktivismus	17
2.3.1.	Darstellung einer Auswahl aktueller Theorien	17
2.3.2.	Kritik der dichotomen Positionierung der aktuellen Theorien nach den Etiketten „konstruktivistisch“ und „instruktionsbestimmt“ (bzw. traditionell, lehrerzentriert, sachorientiert, o.ä.)	19
2.3.3.	Konstruktivismus in Didaktik und Unterricht, ein Dilemma	22
2.3.4.	Kritik der Lerntheorie des Konstruktivismus	23
2.3.5.	Die gemäßigt konstruktivistische Position	24
II.	UNTERRICHTLICHE UMSETZUNG	28
1.	Voraussetzungen	28
1.1.	Fachdidaktische Charakterisierung	28
1.2.	Lernen von „Ökologie“ im Museum	31
1.2.1.	Überlegungen zum Museumsbesuch einer Schulklasse	31
1.2.2.	Museum	31
1.2.3.	Fachliche Charakterisierung	39
1.2.4.	Lernen ökologischer Inhalte	55
2.	Zum Unterricht	57
2.1.	Die Aufgaben der Schüler	57
2.2.	Treatmentcharakterisierung nach Kriterien des gemäßigten Konstruktivismus	58
2.3.	Treatmentcharakterisierung nach Kriterien instruktional orientierten Unterrichts	61
2.3.1.	Vorbemerkung	61
2.3.2.	Die Lernziele (Grobziele, Feinziele) der Treatments	63
2.3.3.	Formulierung, Sprache und Didaktische Reduktion	71
2.3.4.	Überlegungen zum „Erwartungshorizont“	72
2.3.5.	Formale und inhaltliche Charakterisierung der Treatmentaufgaben	73
2.3.5.1.	Formale Charakterisierung	73
2.3.5.2.	Inhaltliche Charakterisierung	73

III.	ENTWICKLUNG DER HYPOTHESEN	76
1.	Hypothesen zur kognitiven Ebene	76
2.	Hypothesen zu beiden Ebenen (kognitiv / affektiv)	77
3.	Hypothesen zur affektiven Ebene	77
IV.	MATERIAL UND METHODEN	80
1.	Gütekriterien der klassischen Testtheorie	80
2.	Vorstudien	83
3.	Stichprobe	85
4.	Erhebungsinstrument / Fragebogen	88
4.1.	Formale Beschreibung	88
4.2.	Inhaltliche Charakterisierung	91
4.2.1.	Konstruktion des Fragebogens	91
4.2.2.	Inhaltliche Analyse des Fragebogens	97
4.3.	Untersuchungstechnische Betrachtung	106
4.3.1.	Untersuchungstechnische Analyse des Fragebogens	106
4.3.2.	Auswertungsobjektivität	109
4.3.3.	Testverfälschungen	112
5.	Auswertung, statistische Behandlung und Darstellung der Daten	115
6.	Untersuchungsdesign	119
7.	Versuchsablauf	120
V.	ERGEBNISSE	124
1.	Befunde zur kognitiven Aussageebene	124
1.1.	Resultate der Gesamtgruppe	124
1.1.1.	Items mit Antwortvorgabe der kognitiven Aussageebene	124
1.1.1.1.	Kognitive Veränderungen in der Gesamtgruppe für die Items mit Antwortvorgabe	124
1.1.1.2.	Vergleich zwischen den Treatmentgruppen	126
1.1.1.3.	Kognitive Veränderungen der Treatmentgruppen	127
1.1.1.4.	Kognitive Veränderungen bezogen auf die Itemtypen	128
1.1.2.	Items ohne Antwortvorgabe der kognitiven Aussageebene	132
1.1.2.1.	Kognitive Veränderungen in der Gesamtgruppe für Items ohne Antwortvorgabe	134
1.1.2.2.	Vergleich zwischen den Treatmentgruppen	135
1.1.2.3.	Kognitive Veränderungen der Treatmentgruppen	136
1.1.3.	Gegenüberstellung der Hauptergebnisse der Items mit Antwortvorgabe (<i>KogMit18</i>) und der Items ohne Antwortvorgabe (<i>KogO</i>)	137
1.2.	Resultate von Teilgruppen	137
1.2.1.	Bekanntheit des Lindenhofs	137
1.2.1.1.	Items mit Antwortvorgabe (<i>KogMit18</i>)	139

1.2.1.1.1.	Ergebnisse der Lindenhof-Nicht-Kenner für Items mit Antwortvorgabe (<i>KogMit18</i>)	140
1.2.1.1.2.	Ergebnisse der Lindenhof-Kenner für Items mit Antwortvorgabe (<i>KogMit18</i>)	142
1.2.1.2.	Items ohne Antwortvorgabe (<i>KogO</i>)	147
1.2.1.2.1.	Ergebnisse der Lindenhof-Nicht-Kenner für Items ohne Antwortvorgabe (<i>KogO</i>)	148
1.2.1.2.2.	Ergebnisse der Lindenhof-Kenner für Items ohne Antwortvorgabe (<i>KogO</i>)	149
1.2.1.3.	Gegenüberstellung der Hauptergebnisse der Items mit Antwortvorgabe (<i>KogMit18</i>) und der Items ohne Antwortvorgabe (<i>KogO</i>) differenziert nach der Bekanntheit des Lindenhofs	150
1.2.2.	Geschlechtsspezifische kognitive Unterschiede	153
1.2.2.1.	Geschlechtsspezifische kognitive Unterschiede von <i>KogMit18</i>	153
1.2.2.2.	Geschlechtsspezifische kognitive Unterschiede von <i>KogO</i>	156
1.3.	Die Befunde zur kognitiven Aussageebene im Lichte ihrer Effektstärken	159
2.	Befunde zur affektiven Aussageebene	160
2.1.	Resultate der Gesamtgruppe	161
2.1.1.	Items mit Antwortvorgabe der affektiven Aussageebene	161
2.1.1.1.	Schülereinschätzung des Museumsbesuchs	162
2.1.1.2.	Vergleich zwischen den Treatmentgruppen	163
2.1.2.	Items ohne Antwortvorgabe der affektiven Aussageebene	164
2.1.2.1.	Sprachliche Äußerungen	164
2.1.2.1.1.	Item zur freien Bemerkung (Fragebogenteil <i>Freie Äußerung</i>)	164
2.1.2.1.2.	Affektive Ebene der Schülerantworten zu Fragebogenteil <i>KogO</i>	164
2.1.2.2.	Zeichnerische Äußerungen	165
2.1.2.3.	Sprachliche und zeichnerische Äußerungen	166
2.2.	Resultate von Teilgruppen	167
2.2.1.	Bekanntheit des Lindenhofs	167
2.2.1.1.	Items mit Antwortvorgabe	167
2.2.1.2.	Items ohne Antwortvorgabe	169
2.2.2.	Geschlechtsspezifische affektive Unterschiede	171
2.2.2.1.	Items mit Antwortvorgabe	171
2.2.2.2.	Items ohne Antwortvorgabe	173
2.3.	Die Befunde zur affektiven Aussageebene im Lichte ihrer Effektstärken (beschränkt auf die Testebene der Items mit Antwortvorgaben <i>AffMu</i>)	174
3.	Kognitive und affektive Ergebnisse zum Kindermuseum	176

VI.	DISKUSSION	178
1.	Überlegungen zu Theorie und Aussagebereich	178
2.	Die zentralen Anliegen der Untersuchung	180
3.	Diskussion methodischer Aspekte	180
4.	Bedeutung der kognitiven Ergebnisse	186
4.1.	Items mit Antwortvorgabe (<i>KogMit</i>)	187
4.2.	Items ohne Antwortvorgabe (<i>KogO</i>)	198
4.3.	Die Effektstärke von Museumsbesuch und Treatment	200
4.4.	Der Einfluss der Bekanntheit des Lindenhofs	201
4.5.	Geschlechtsspezifische Phänomene	203
5.	Bedeutung der affektiven Dimension	205
6.	Unterrichtliche Konsequenzen	207
VII.	ZUSAMMENFASSUNG	209
VIII.	LITERATUR	212
IX.	ANHANG	

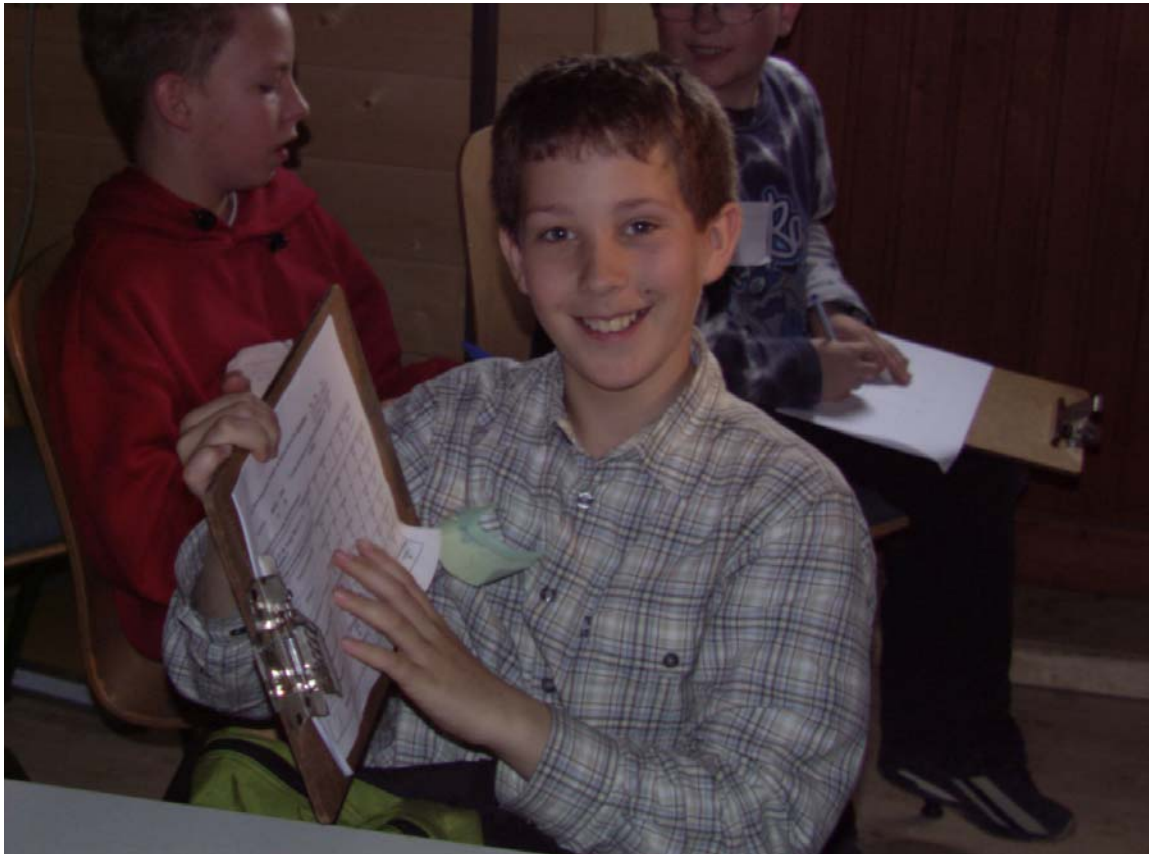


Abbildung 1: Ein Schüler bearbeitet Im Museum des Lindenhofs seinen „NT I“.

Abkürzungsverzeichnis:

S	vorwiegend selbstbestimmt
SF	Mischform zwischen selbst- und fremdbestimmt
F	weitgehend fremdbestimmt
	Diese Abkürzungen bezeichnen zuerst die entsprechenden Treatments, werden in der Folge jedoch auf entsprechende Teilgruppen, Einzelschüler, Probanden, Jungen, Mädchen, etc. bezogen.
S-Item	auf S-Treatment bezogenes Item; S-Schüler sollten diese Items besonders gut beantworten können.
SF-Item	auf SF-Treatment bezogenes Item; SF-Schüler sollten diese Items besonders gut beantworten können.
F-Item	auf F-Treatment bezogenes Item; F-Schüler sollten diese Items besonders gut beantworten können.

<i>KogMit</i>	Testteil aus kognitiven Items mit Antwortvorgaben (meist ist der Testteil <i>KogMit18</i> gemeint)
<i>KogMit18</i>	Testteil aus kognitiven Items mit Antwortvorgaben zu den zentralen 18 Items zur Bewertung des kognitiven Lernerfolgs
<i>KogMit25</i>	Testteil aus kognitiven Items mit Antwortvorgaben zur erweiterten Itemauswahl (z.B. incl. Items zum Kindermuseum)
<i>KogO</i>	Testteil aus kognitiven Items ohne Antwortvorgaben
<i>AffMu</i>	Testteil aus Items zur affektiven Ebene zu Museumsbesuch und Stationen
<i>AffMuUr</i>	Testteil aus Items zur affektiven Ebene zu Einzelelementen der Station Urwald
<i>AffMuPla</i>	Testteil aus Items zur affektiven Ebene zu Einzelelementen der Station Platzbedarf
H ₁ bis H ₅	Hypothesen eins bis fünf
VT	Vortest
NT I	Nachtest unmittelbar nach dem Museumsbesuch
NT II	Nachtest 40 Tage nach dem Museumsbesuch
Tr	Treatment
Kgr	Kleingruppen
St	Station
W	Wechsel
Aph	Arbeitsphase
Ki	Kindermuseum
p	kleine Pause
P	große Pause
U	Untersuchungstag
B	Beobachtungstag

L	Lehrer
min	Minute
B1	Beurteilungswert 1 (nach Bloom 2001)
B2	Beurteilungswert 2 (nach Bloom 2001)
Rp	Reproduktion
Ro	Reorganisation
T	Transfer
P	Problemlösen
CLT	Cognitive load theory nach Owen & Sweller (1985)
r	Korrelationskoeffizient
N	Anzahl
n.s.	nicht signifikant
sig.	signifikant
p	Wahrscheinlichkeit
TSK	korrigierter Trennschärfekoeffizient
SI	Schwierigkeitsindex

I. Einleitung

1. Zum Forschungsgegenstand

Nach dem „PISA-Schock“ (z.B. Die Zeit 50/2003: Die Bilanz des Schreckens; vgl. Baumert et al. 2002; vgl. <http://www.pisa.oecd.org/> am 21.05.2004), einer weitgehenden Entrüstung der Öffentlichkeit in Deutschland über die schwachen Leistungen deutscher Schüler im internationalen Vergleich in einer von der OECD durchgeführten Untersuchung (PISA: „Programme for International Student Assessment“), hat man in Deutschland allen Grund, sich mit Methoden auseinanderzusetzen, die bessere Bildung hervorbringen können. Das gilt auch für das Fach Biologie, selbst wenn laut „Erhebungszyklus“ der Teil der PISA-Studie mit dem Schwerpunkt „Naturwissenschaftliche Grundbildung“ erst im Jahre 2006 (Baumert et al. 2002, S. 13) stattfinden wird. Die Ergebnisse der TIMS-Studie, „Third International Mathematics and Science Study“, aus dem Jahre 1997, die sich mit mathematisch-naturwissenschaftlichem Unterricht im internationalen Vergleich befasste, sind beunruhigend genug (Baumert et al. 1997, S. 55): „Die deutschen Schülerinnen und Schüler erreichen das Leistungsniveau der internationalen Mittelgruppe in einem im Durchschnitt um sechs bis zwölf Monate höheren Lebensalter als die Schülerinnen und Schüler aller anderen Länder dieser Gruppe.“

Im Kontext der PISA-Ergebnisse sieht Horster (2002, S. 9 ff.) die Hauptschwäche deutscher Schüler in der Anwendung von fachspezifisch erworbenem Wissen. Nach Mayer (2002, S. 82) ist für die schlechten Ergebnisse in den PISA-Tests v. a. „die Ebene des Unterrichts“ verantwortlich. Der Unterricht ist weiterzuentwickeln „(...) in Richtung auf ein erweitertes Verständnis von Lernen“ (Horster 2002, S. 11). Horster lässt keinen Zweifel an den für ihn bedeutsamen Merkmalen anzustrebender Lernsituationen, wie z.B. sinnstiftende Kontexte, authentische Situationen und schüleraktivierende Methoden; er strebt zumindest die Integration von Phasen situierten Lernens in den ‚real existierenden‘ Unterricht an (Horster 2002, S. 12 f.). Kurz: Horster fordert, auch wenn er dies an keinem Punkt explizit formuliert, an konstruktivistischen Leitgedanken orientiertes Lernen.

Der Konstruktivismus als Lerntheorie hat in den letzten Jahrzehnten stark an Bedeutung gewonnen. Schulz-Zander (1997, S. 269) spricht von einem „lerntheoretischen Paradigmenwechsel“ von traditionellen, noch behavioristisch geprägten Konzepten hin zum Lernen „im konstruktivistischen Sinne“. Tobin (1993) beobachtet „constructivism has become increasingly popular.“ Matthews (1998, S. 1) attestiert dem Konstruktivismus „enormous influence (...) on contemporary science education“ und deutet durch seine Zusammenfassung die Breite des Einflusses an: „The article distinguishes educational constructivism (...), from constructivism in the philosophy of science (...) and from constructivism in the sociology of science (...). It notes the expansion of educational constructivism from initial considerations of how children come to learn, to views about epistemology, educational theory, ethics, and the cognitive claims of science“. Der Umfang an Arbeiten zu konstruktivistischen Ansätzen ist enorm (vgl. Einleitung 2.2.). Wie dagegen die Ergebnisse der TIMS-Studie belegen (Baumert et al. 1997, S. 215 und 231 f.), ist zumindest Mathematikunterricht in Deutschland „Wissenserwerbunterricht“, der klar vom Lehrer dominiert wird. Man kann daher vermuten, dass diese „neue Lerntheorie“ bisher nicht in den Schulen angekommen ist (vgl. Gruber et al. 2000, S. 140).

Wichtige von konstruktivistischen Lerntheorien geforderte Lernbedingungen sind Authentizität und Situiertheit. Diese Voraussetzungen erfüllt Unterricht in außerschulischen Lernorten, wie in Naturkundemuseen, meist besser als Unterricht im Klassenraum. Die Literatur bietet keine überzeugenden empirischen Befunde, ob eine Umsetzung von konstruktivistischen Maximen bei einem Unterrichtsbesuch im außerschulischen Lernort „Naturkundemuseum“ dem Lernerfolg zuträglich ist. Wenige naturwissenschaftlicher Fachdidaktik zuzuordnende Artikel befassen sich mit Naturkundemuseen. Zu nennen sind z.B. Gries (1996) und Frank (2001). Gries beschäftigt sich auf sehr allgemeiner Ebene mit dem Unterricht im Naturkundemuseum. Von konkreter empirischer Überprüfung berichtet sie nicht. Frank beschreibt den Forschungsstand in der Geografiedidaktik zum Unterricht im Museum und beklagt die wenig fundierten und nicht evaluierten Befunde (Frank, 2001, S. 21). Seine Untersuchung bestätigt, dass unterrichtliche Bedingungen, wie z.B. Methodenwechsel und Lernzielkontrollen, den Lernerfolg nicht nur im Klassenzimmer, sondern auch im Museum erhöhen (Frank, 2001, S. 131 ff.).

Die obige Ausgangsfrage, inwieweit konstruktivistisch orientierter Unterricht gegenüber traditionellem Unterricht zu besserem Lernerfolg führt, kann man mit Hilfe dieser Arbeiten nicht beantworten.

Ein außerschulischer Lernort kann Gelegenheit für den Einsatz methodischer Alternativen zum Unterricht im Klassenraum geben (vgl. Killermann 1999, S. 11). Durch den Museumsbesuch wird das Beharren der Schüler auf den gewohnten Unterricht vermindert und es ist möglich, in dieser Quasi-Laborsituation - ein für die Durchführung einer Untersuchung zusätzlich günstiger Umstand - Neues zu erproben. Zudem ist nach Becker (2000, S. 24) ein außerschulischer Rahmen zur Vermittlung ökologischer Inhalte, und v. a. darum geht es in dem hier untersuchten Naturkundemuseum, geeigneter als die Schule. Naturkundemuseen sind „Erlebnis- und Lernorte“ (Gries 1996, S. 8), beziehen somit affektive Ebenen der Rezeption ein.

In Meta-Analysen, die sich auf 150 Studien zu Formen „offenen Unterrichts“ im Vergleich zu „traditionellem Unterricht“ beschäftigen, zeigt sich der Unterschied dieser beiden Formen in ihrer kognitiven und affektiven Wirkung auf Schüler als so gering, dass diese Differenz als „pädagogisch (...) nicht bedeutsam“ eingestuft wird (weniger als ein Prozent erklärter Varianz; Häußler et al. 1998, S. 152 ff.). Sicher ist Vorsicht geboten bei einer Übertragung von Ergebnissen zum „offenen Unterricht“ auf Lernformen nach konstruktivistischen Maximen. Dennoch zeigt dies, wie wenig effizient bisher erprobte vergleichbare Unterrichtsverfahren sind.

Konstruktivismus, konstruktivistische Vorstellungen und sich daraus ableitende Befunde für Lernsituationen werden kontrovers diskutiert (z.B. Tobin 1993, Matthews 1998, Suchting 1998, Glaserfeld 1998, Hoops 1998, Terhart 1999). Auf welcher Ebene dieser Konflikt z.T. ausgetragen wird, sei kurz referiert: Matthews (1998, S. 2) zitiert Tobin (1993, Preface): „there is widespread acceptance of constructivism.“ Fischler (1999, S. 576) folgt in seiner Rezension über Matthews „Constructivism in Science Education“ Matthews Kritik der „unbegründeten Selbstsicherheit und Überheblichkeit bei den Konstruktivisten“, „etwa bei der Behauptung von Tobin, der Konstruktivismus sei verbreitet akzeptiert.“ Hier das vollständige Zitat Tobins (1993, Preface): „Nevertheless, the revolution has progressed steadily (gemeint ist ein Aufbegehren gegen „traditional approaches to educational reform“) and there is wi-

despread acceptance of alternatives to objectivism, one of which is constructivism, the focus of this book.“ – Man kann sich bei der Verkürzung von Tobins Zitat durch Matthews – interessanterweise wird genau diese Aussage von Fischler aufgenommen – des Eindrucks einer interessegeleiteten Unkorrektheit kaum erwehren. Tobin (1993, Preface) verstärkt diesen Eindruck: „Currently there is a paradigm war raging in education.“

Weiter zitiert Matthews (1998, S. 1 f.) heftige Kritik von Devitt (1991): „Constructivism attacks the immune system that saves us from silliness.“ In dieses Bild passt auch der von Fischler (1999, S. 576) beschriebene „Glaubenskampf“ zwischen Suchting (1998) und Glaserfeld (1998), „in dem Argumente (...) mit großer Schärfe vorgebracht [werden]“. Weit weniger polemisch argumentieren Terhart (1999) und Hoops (1998) (vgl. unten).

Die zentrale Frage dieser Untersuchung lautet: Was fördert unterrichtlichen Erfolg besser, instruktional orientierter oder konstruktivistisch orientierter Unterricht? Unterrichtserfolg sollte bezogen auf Thema (Ökologische Inhalte) und Ort (Naturkundemuseum) der Untersuchung nicht ausschließlich kognitive Ebenen umfassen, sondern die affektive Wirkung der Interventionen einbeziehen. Die zu konzipierende Unterrichtsstudie soll dem von Krüger (2003; vgl. Fischer et al. 2003) formulierten biologiepädagogischen „Forschungsrahmen“ für „entwicklungsorientierte Evaluationsforschung“ genügen: Auf der Basis einer Theorie, wie z.B. der des Konstruktivismus, ist ein Lernangebot zu entwickeln, welches in einer hypothesengeleiteten empirischen Untersuchung zu überprüfen ist. Verwiesen sei auf Kapitel III. Entwicklung der Hypothesen.

2. Lerntheoretische Grundlagen: Konstruktion und Instruktion

Angesichts der oben geschilderten „Grabenkämpfe“ erscheint vor der Entwicklung von Unterrichtsinterventionen anhand konstruktivistischer Ansätze und deren empirischer Überprüfung eine kritische theoretische Auseinandersetzung mit dem Konstruktivismus unabdingbar.

In dieser Untersuchung werden unterschiedliche Treatments hinsichtlich ihrer kognitiven und affektiven Wirkung auf Schüler miteinander verglichen. Die Treatments beziehen sich auf die **Lerntheorie des Konstruktivismus** und die Gegenposition **instruktionsbestimmter Lehr-Lern-Prozesse**. Diese dichotome Positionierung der beiden Ansätze beschreibt Hoops (1998, S. 235 ff.), verwenden Reinmann-Rothmeier & Mandl (2001) und findet sich in ähnlicher Form bei Issing (1997, S. 197).

Reinmann-Rothmeier & Mandl (2001) stellen die Dichotomie von Konstruktion und Instruktion sehr akzentuiert dar: „Konstruktivistisch gefärbte Auffassungen“, Lernprozesse, in denen das „Primat der Konstruktion“ vorherrscht, zeichnen sich durch eine „aktive Position des Lernenden“ aus; Lernen ist ein „konstruktiver, situativer Prozess“; Unterrichten heißt „Unterstützen, Anregen, Beraten“ (Rothmeier & Mandl (2001, S. 604 und 614 ff.). Die Rolle des Lehrenden ist im Sinne Leinhardts (1993, S. 3 ff.) die eines „facilitators“, zu verstehen im Sinne von „Ermöglicher“.

Die Gegenposition charakterisieren Reinmann-Rothmeier & Mandl (2001, S. 604 ff.) als „Gegenstandszentrierte Lernumgebungen“, in welchen das „Primat der Instruktion“ vorherrscht. Unterricht ist demzufolge „Anleiten, Darbieten, Erklären“. Der Lehrende ist aktiv, der Lernende rezipiert passiv. Das Ziel ist es, den Gegenstand des Lehr-Lern-Prozesses so weiterzugeben, dass der Lerner ihn am Ende dieser Wissensübergabe genauso besitzt wie der Lehrende (Reinmann-Rothmeier & Mandl 2001, S. 606). Skinner (1971, S. 11) pointiert diese Position: „Er (der Lehrer) gibt, der Schüler nimmt.“

Im Folgenden wird die Lerntheorie des Konstruktivismus im Blick auf die „Gegentheorie“ instruktionsbestimmter Lehr-Lern-Prozesse im Kontext ihrer theoretischen und historischen Bezüge vorgestellt. Aktuelle Spielarten der Lerntheorie des Konstruktivismus werden umrissen und aus einer dieser Theorien, „einer gemäßigt konstruktivistischen Position“ (Reinmann-Rothmeier & Mandl 2001, S. 626 ff.), leiten sich Prämissen zur Gestaltung der Treatments dieser Untersuchung ab. Die Treatments werden im Sinne konstruktivistischer Kriterien und nach „traditioneller“, d.h. instruktional orientierter, Vorgehensweise beschrieben.

2.1. Aktueller Theoriekontext der Lerntheorien

2.1.1. Konstruktivismus

Der Konstruktivismus hat viele Spielarten. Die Verwendung des Begriffs zur Bezeichnung einer Kunstrichtung im beginnenden 20. Jahrhundert (vgl. Watzlawick 2001, S. 9; Lodder 1985; Winkler 1981) wird nicht in die folgende Analyse einbezogen. Das gilt auch für Begriffe in der *ursprünglichen* Wortbedeutung, die sich nicht auf konstruktivistische Theorien beziehen, wie z.B. bei Klafki (1980) „Zur Unterrichtsplanung im Sinne kritisch-konstruktiver Didaktik“ oder bei König, E. und Riedel (1971) „Unterrichtsplanung als Konstruktion“.

Für eine theoretische Betrachtung des Konstruktivismus aus einer Meta-Perspektive heraus bieten Gerstenmaier & Mandl (1995) und Terhart (1999) sehr brauchbare Strukturierungshilfen zur Orientierung in der Vielfalt konstruktivistischer Strömungen an. Gerstenmaier & Mandl (1995) lassen eine Auffächerung in drei Hauptbereiche erkennen (vgl. ähnliche Sichtweise Matthews 1998):

- Konstruktivismus als Erkenntnis- und Wissenschaftstheorie
- „Neuer“ Konstruktivismus in der Soziologie, Kognitionswissenschaft und Psychologie
- Konstruktivistische Ansätze in der Instruktionspsychologie und empirischen Pädagogik

Wie schon diese Bezeichnungen andeuten, fächert sich jede Strömung weiter auf, so dass ein komplexes Gesamtbild entsteht.

Terhart (1999, S.631) dagegen unterscheidet vier hauptsächliche Theoriekontexte konstruktivistischer Didaktik. Damit beschränkt er seine Analyse von vornherein, liefert dadurch jedoch einen Überblick, der für die vorliegende Untersuchung angemessen erscheint. Die Theoriekontexte sind (vgl. Abbildung 2):

- Radikaler Konstruktivismus
- Neurobiologie des Erkennens
- Systemtheorien
- Kognitionspsychologische Lernkonzeptionen

Diese einfachere Strukturierung in vier von Terhart nicht weiter differenzierte Bereiche soll als Ausgangspunkt zur Hinführung und Bereitung des theoretischen Problemgrundes konstruktivistischer Positionen in der Lehr-Lern-Forschung genügen.

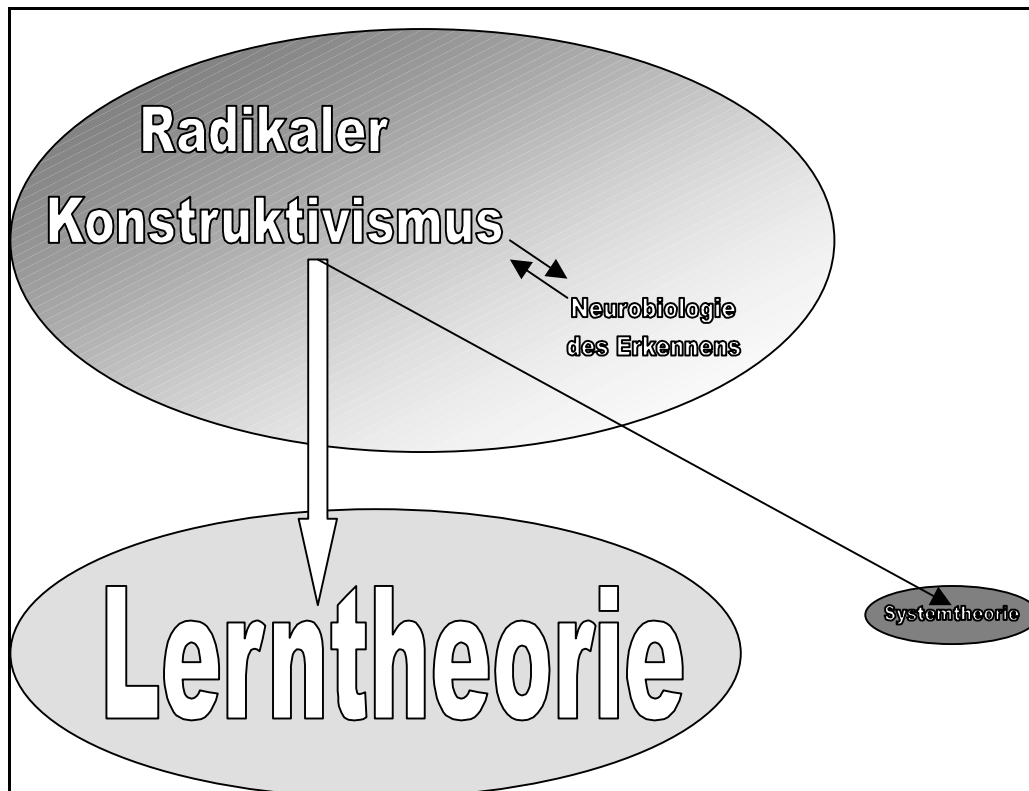


Abbildung 2: Vier Theoriekontexte konstruktivistischer Didaktik (Terhart 1999). Die Anordnung und Schriftgröße deuten Interpretation und Bedeutung für diese Untersuchung an.

Zuerst sei die **Erkenntnistheorie des „Radikalen Konstruktivismus“** (Glaserfeld 1987) zusammen mit der **„Neurobiologie des Erkennens“** ausgeführt. „Die ‚Neurobiologie des Erkennens‘ wird herangezogen, um erkenntnistheoretische Annahmen empirisch zu begründen“ (Terhart 1999, S. 632). Nach Terhart (1999, S. 633) deuten „führende Neurobiologen (...) ihre Ergebnisse auf der Theorie-Ebene [selbst] radikal-

konstruktivistisch“. Damit lässt sich eine eigenständige Theorie der „Neurobiologie des Erkennens“ nicht entdecken, vielmehr wird ein Gedankengebäude (Radikaler Konstruktivismus) empirisch gestützt (neurobiologische Befunde). Eine eigene Behandlung der „Neurobiologie des Erkennens“ wird deshalb für entbehrlich gehalten.

Kerngedanken des Radikalen Konstruktivismus sind in folgendem Zitat gebündelt: „Die Welt, in der wir zu leben meinen, [haben] wir uns selbst zu verdanken“ (Glaserfeld 2001, S. 17).

Als erstes bedeutet diese zentrale Erkenntnis des Radikalen Konstruktivismus, „alle Verständigung, alles Lernen und Verstehen [ist] Bau und Interpretation des erlebenden Subjekts (...)“ (Glaserfeld 2001, S. 17). Wir, als Subjekte, bauen unsere Welt selbst zusammen. Wir konstruieren unsere Wirklichkeit (vgl. Watzlawick 2001, S. 9)! Roth (1997, S. 21) formuliert: „Die Welt, in der ich lebe, ist ein Konstrukt des Gehirns.“ „Wirklichkeit [ist] neuronal konstruiert“ (Terhart 1999, S. 633, vgl. Friedrich & Preiss 2002, S. 64). Ramachandran (1998, S. 380) führt für bestimmte visuell vermittelte sehr schnelle Lernprozesse Beispiele stofflicher Veränderung bestimmter Neuronen im Gehirn auf. Tovee et al. (1996) weisen Entsprechendes in gehirnphysiologischen Untersuchungen an Makaken nach und vermuten analoge Phänomene beim Menschen. Es finden demnach zeitgleich zu Konstruktionsprozessen auf der Ebene des subjektiven Verstehens Veränderungen im Gehirn auf stofflicher Ebene statt (Näheres siehe Tovee et al. 1996), so dass man sich eine Verknüpfung zwischen den beiden Ebenen vorstellen kann. Zu beachten bleibt, die Ebenen nicht zu vermengen. Man befindet sich auf der Aussageebene einer Korrelation, nicht der einer Kausalität. Unter diesem Aspekt sind die obigen Aussagen von Roth und Terhart zu relativieren.

Das obige Zitat hat eine weitere Bedeutung. Glaserfeld (2001, S. 17) geht davon aus, „dass wir die Operationen, mit denen wir unsere Erlebniswelt zusammenstellen, weitgehend erschließen können“. Damit haben wir ein gewisses Maß an Kontrolle über diese konstruierte Welt und so auch Verantwortung für unser Denken und Handeln (Glaserfeld 1998, S. 125). Hier ist einzuschränken, dass alle Erkenntnisse über Operationen, mit denen wir unsere Welt zusammenbauen, selbst zu diesen Operationen gehören (vgl. Abbildung 3).

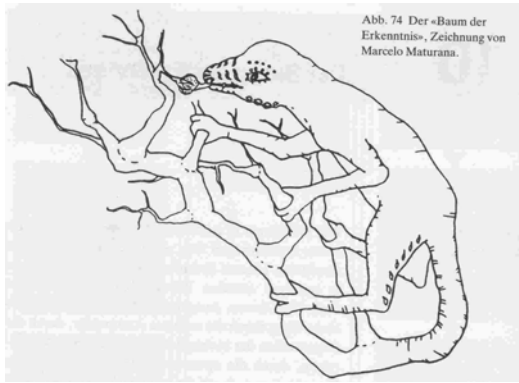


Abbildung 3: Der „Baum der Erkenntnis“ (Maturana & Varela 1987).

Dennoch ist Konstruktivismus kein Solipsismus¹ (Terhart 1999, S. 631). Es wird nicht die Existenz einer Realität außerhalb des Subjekts geleugnet (Mitterer 1998, S. 111), sondern ihre objektive Erkennbarkeit durch das Subjekt bzw. den menschlichen Verstand. „(...) all das, was wir Wissen nennen [ist] (...) von den Wahrnehmungs- und Denkmöglichkeiten des erkennenden Menschen bestimmt“ (Glaserfeld² 1996, S. 101, vgl. Foerster 2001). „(...) alle meine geistigen Leistungen (...) [sind] Leistungen meines Gehirns“ und „unterliegen (...) den biologischen Konstruktions- und Funktionsbedingungen meines Gehirns“ (Roth 1997, S. 22). Vollständiges Erkennen – das gilt auch für jede Metaebene des Erkennens –, Allgemeingültigkeit jeglicher mit menschlichen Mitteln erkannter Regel ist per se nicht möglich. Roth (1997, S. 23): „Ich glaube ja auch nicht, dass die Konstruktionen eines Ameisengehirns die objektive Wahrheit widerspiegeln, warum dann diejenigen des Menschen?“ Glaserfeld (2001, S. 19) richtet diese konstruktivistische Perspektive auch auf ihn selbst betreffende Domänen, nämlich Wissenschaft und Philosophie: „Die meisten Wissenschaftler [fühlen] sich auch heute noch als ‚Entdecker‘, die Geheimnisse der Natur lüften und den Wissensbereich langsam aber sicher erweitern; und unzählige Philosophen widmen sich der Aufgabe diesem mühsam errungenen Wissen die unumstößliche Sicherheit zuzuschreiben, die alle Welt von der ‚echten‘ Wahrheit erwartet. Nach wie vor herrscht die Auffassung, dass Wissen nur dann Wissen ist, wenn es die Welt erkennt wie sie ist“. Dieses objektive Erkennen lehnt der Radikale Konstruktivismus ab!

¹ Wenngleich Piaget (1974), der als früher Verfechter konstruktivistischer Gedanken gilt, dem Neugeborenen eine solipsistische Perspektive zubilligt, die im Wechselspiel von Assimilation und Akkommodation durchbrochen wird.

² Glaserfeld bezieht sich hier auf Kant.

Fischler (1999, S. 572) äußert sich ähnlich: „Vermutlich sind die meisten Naturwissenschaftler in ihrem Innersten Realisten insofern, als sie überzeugt sind, dass sie mit ihren Forschungsergebnissen die Wirklichkeit ganz gut beschreiben. Konstruktivisten sind weit entfernt von dieser Annahme (...)“. Konstruktivisten, insbesondere radikale Konstruktivisten wie Watzlawick (2001, S. 38), richten ihre Sichtweise auch auf die Wissenschaft: „Die Wissenschaft konstruiert Weltbilder (...)“.

Können wir dann unsere subjektive Welt konstruieren, wie es uns beliebt? Nach Glaserfeld (2001, S. 19) sind unsere Konstruktionen „Anpassungen im funktionalen Sinn³“. Analog zur Evolutionstheorie formuliert Glaserfeld (2001, S. 19) ein Wechselspiel zwischen der Konstruktion jedes Menschen und der Realität⁴. Ist eine Konstruktion den Erfordernissen der Realität nicht angepasst, so wird sie sich im Wettbewerb der Konstruktionen nicht bewähren. Ein entscheidender Unterschied zur Evolutionstheorie liegt darin, dass die Konstruktionen eines Subjekts auf der Ebene erkenntnistheoretischer Diskussion nicht als biologisches Merkmal aufzufassen sind, das die Fitness (z.B. im Sinne Immelmanns 1982 oder Wilsons 1975) eines Individuums beeinflusst. Das Subjekt misst seine Konstrukte an der Umwelt und verändert sie entsprechend. Unsere Konstrukte müssen in unsere materielle und soziale Umwelt passen, d.h. viabel sein (Hansmann 1998, S. 89). „Constructivism replaces the notion of truth with that of viability (...)“ (Glaserfeld 1993, S. 27).

Die sich aus dem Konstruktivismus ableitende Systemtheorie, „**Theorie autopoietischer selbstreferentieller Systeme**“ (Terhart 1999, S. 633 f.), sei nur erwähnt, da eine Relevanz für Lernprozesse sich nicht unmittelbar erschließt. Alle als Systeme vorstellbaren Seinsweisen (Entitäten) - das gilt z.B. für Lebewesen; Maturana & Varela (1987, S. 55) betonen jedoch, dass Lebewesen nicht die einzigen autonomen, zur Autopoiese befähigten „Wesen“ sind - erschaffen sich selbst, sind selbstreferentiell und stehen begrenzt im Austausch mit der Umwelt. Sie werden von anderen Systemen wahrgenommen und nehmen selbst andere Systeme wahr (Terhart 1999, S. 633 f.).

³ Ein Gedanke, den Glaserfeld von Piaget übernimmt (Glaserfeld 1996, S. 102 und 1998, S. 124).

⁴ Verwendet im Sinne der Termini von Roth (1997, S. 324). Er differenziert zwischen Wirklichkeit, als die von unseren Gehirnen hervorgebrachte Konstruktion von Welt, und Realität, als die nicht objektiv erkennbare Welt, in der das Gehirn als Konstrukteur existiert.

2.1.2. „Instruktivismus“

Die Darstellung eines Gegenentwurfs, der sich auf instruktionsbestimmte Lehr-Lern-Prozesse bezieht, stellt sich als nicht unproblematisch dar. Obwohl einzelne Vertreter instruktionsbestimmten Unterrichtens theoriegeleitet vorgehen, wie z.B. Skinner (1971), der seine Vorstellungen zum Programmierten Unterricht aus behavioristischen Theoriekonzepten entwickelt, so gibt es keine gemeinsame stringente „Theorie der Instruktion“, etwas wie „Instruktivismus“; es gibt sie nicht als Lerntheorie und es gibt sie nicht als Erkenntnistheorie. Dennoch haben die als instruktionsbestimmt identifizierten Lehr-Lern-Theorien gemeinsame Merkmale, wie oben aus Reinmann-Rothmeier & Mandl (2001) zitiert.

Aus diesen Merkmalen kann man versuchen, passende erkenntnistheoretische Grundlagen abzuleiten. Bei diesen instruktionsbestimmten Lehr-Lern-Vorgängen geht man von der Möglichkeit einer im Ideal verlustfreien und nicht-modifizierenden Weitergabe von Wissen vom Lehrer zum Lerner aus. Gut vorstellbar bei dieser Perspektive ist, dieses Wissen als interne Repräsentation der objektiven Realität zu betrachten. Dies würde den Gedanken identischer interner Repräsentationen unterschiedlicher Individuen vereinfachen. Dann nämlich wäre ein „Eichen“ dieser Repräsentationen an der Realität möglich. Das entspricht folgendes von Watzlawick (2001, S. 9) artikulierten Gedanken: „Was wir wissen, gilt im allgemeinen für das Ergebnis unserer Erforschung der Wirklichkeit. Von dieser Wirklichkeit nimmt der gesunde Menschenverstand an, dass sie gefunden werden kann.“ Man geht von der gänzlich unkonstruktivistischen Annahme aus, die erkennbare Wirklichkeit sei ein genaues Abbild der Realität bzw. könne sie sein. Demzufolge wäre der instruktionsbestimmten Lehr-Lern-Theorie die dem Konstruktivismus konträre Erkenntnistheorie des Objektivismus zuzuordnen, nach der „dem Erkenntnisvermögen [eines Menschen] die Fähigkeit zugeschrieben wird, vom Subjekt unabhängige reale Gegenstände und objektive Ideen zu erfassen“ (Brockhaus 1998). Im Folgenden wird deshalb der Begriff „Instruktivismus“ im o. g. Sinne verstanden und verwendet.

2.1.3. Fazit

Die erkenntnistheoretischen Grundpositionen zu „Konstruktion“ und „Instruktion“ unterscheiden sich in Bezug auf die Lerntheorien v.a. durch folgende Punkte: Eine „Theorie zur Instruktion“ könnte objektiv erkennbares Wissen zulassen. Die Möglichkeit unveränderter Weitergabe interner Repräsentationen wird angenommen. Im Konstruktivismus dagegen können Subjekte kein objektives Erkennen erreichen. Jedes Wissen ist subjektive Konstruktion und kann nur vom Subjekt selbst aktiv aufgebaut (konstruiert) werden. Genau diese Punkte sind entscheidend für lerntheoretisch relevante Vorstellungen, z.B. den „neuen Konstruktivismus in der Pädagogischen Psychologie [, der sich] nicht mit grundlegenden Prinzipien menschlicher Erkenntnis, sondern mit den Prozessen des Denkens und Lernens handelnder Subjekte [beschäftigt]“ (Reinmann-Rothmeier & Mandl 2001, S. 614 f.).

2.2. Historischer Kontext von Instruktion und Konstruktion

Zum Konstruktivismus als Lerntheorie gibt es in der aktuellen Literatur der Didaktik der Naturwissenschaften und der Mathematik eine Fülle von Veröffentlichungen. Fischler (1999) bespricht dazu acht entsprechende Werke, die alle selbst Aufsatzsammlungen darstellen. Gunstone (2000, S. 254) berichtet von Tausenden von Fundstellen zum Begriff „constructivist classroom“ und „constructivist“ in didaktisch orientierten und anderen Datenbanken und folgert „[this] suggests that there are no areas of human activity to which the label ‚constructivist‘ is not currently being applied in some form.“ Matthews (1998, S. 2) berichtet über Literatur zu „constructivist research“: „A 1990 bibliography produced at Leeds University (...) listed over 1000 works.“ Duit trägt seit 1984 (anfangs zusammen mit Pfundt) eine Bibliographie namens „Students' and Teachers' Conceptions and Science Education“ zusammen, von der er im Vorwort sagt: „This research has been carried out within what is called the constructivist view including individual constructivist and social constructivist perspectives. The bibliography may therefore now be viewed as an attempt to document constructivist research in science education.“ Die Bibliographie enthält weit über 6000 Artikel (Duit 2004). Gleichzeitig werden konstruktivistische Vorstellungen und entsprechende Lernprozesse kontrovers diskutiert (z.B. Glaserfeld 1998, Matthews 1998, Suchting 1998). Im Folgenden werden Vorläufer dieser Lerntheorie bis ins frü-

he 20. Jahrhundert nachgewiesen. Instruktionsbestimmte Vorstellungen von Lehr-Lern-Prozessen bzw. entsprechenden Lerntheorien werden aktuell oft kritisch beurteilt (z.B. Baptist 2003, Klautke 2003, Mandl et al. 1997), scheinen aber dennoch den gegenwärtigen Unterricht in Deutschland, und z.B. auch in den USA zu dominieren. Historische Vorbilder, die bis ins 19. Jahrhundert hinein reichen, werden aufgeführt.

Instruktionsbestimmte Vorstellungen von Unterrichten, „klassischer Frontalunterricht“, mit dem Lehrenden im Mittelpunkt prägen nicht nur ironisierende historische Sichtweisen wie die des Lehrers Lämpel mit strengem Blick und erhobenem Zeigefinger von Wilhelm Busch von 1865 (1959, S. 39) (Abbildung: 5) oder historische Darstellungen in einer „Dorfschule von 1848“ mit einem Lehrer als Dompteur in aggressiver Pose mit Rohrstock inmitten braver Schüler in einem überfüllten Klassenraum von Albert Anker (Abbildung: 4). Eine Vielzahl historischer didaktischer Vorstellungen leitet sich von diesem Bild ab, wie auch Vorstellungen, die bis in die Gegenwart reichen.



Abbildung 4 und 5: Dorfschule und Lehrer Lämpel (vgl. Baptist 2003). Weiteres siehe Text.

2.2.1. Zur Tradition der Instruktion

Reinmann-Rothmeier & Mandl (2001, S. 610 ff.) nennen z.B. die Formalstufen des Unterrichts nach Ziller (1876), die Konzeption des „sinnvoll rezeptiven Lernens“ nach Ausubel (1974), den Programmierten Unterricht nach Skinner (1971) oder behavioristische Instruktionsdesign-Modelle z.B. von Bloom (1976). Planung des Unterrichts, Operationalisierbarkeit der zu erreichenden Ziele, auf kognitiver und z.T. auch auf affektiver Ebene, ist in diesem Zusammenhang ein wichtiges Thema, mit dem sich

viele Abhandlungen beschäftigen, z.B. Mager (1965), NSTA (1969), Krathwohl (1975), Bloom (1976), Möller (1976), Hacker (1979) und Westphalen (1980). Der letztgenannte Autor formuliert seinen „ersten Grundsatz“, der eine „in allen Ländern der Bundesrepublik“ befolgte „Selbstverständlichkeit“ sei, klar: „Jede Unterrichtsplanung muss ausgehen von klar definierten, begründeten Lernzielen“ (Westphalen 1980, S. 16).

Issing (1997, S. 197) rechnet dieses Lernen, bei dem die Vermittlung im Zentrum steht, als dem Instruktionsparadigma zugehörig und nennt ebenfalls Skinner und Ausubel als wichtige Vertreter dieser Position. Dabei ist es keineswegs so, dass instruktional ausgerichtete Vorstellungen von Lernprozessen nur historische Bedeutung haben. Issing (1997) selbst beschreibt systematisches Instruktionsdesign für die Entwicklung von Lernsoftware, wenngleich ihm offensichtlich Kritik und Gegenposition bekannt sind. Die Neubearbeitung von Blooms Lernzieltaxonomie (Anderson et al. 2001) spricht für die aktuelle Relevanz zielorientierter und instruktionsbestimmter Unterrichtsformen.

Diese Art des Unterrichtens ist auch in der Gegenwart im schulischen Alltag weit verbreitet. White & Gunstone (1992, Abbildung 6) zeigen typische Beispiele der Vorstellung von Unterrichten gezeichnet von Studenten zu Beginn ihrer Ausbildung zum Lehrer.

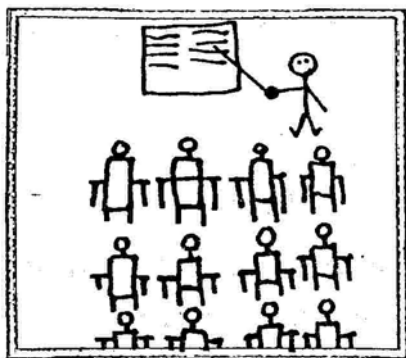


Abbildung 6: Typische Vorstellungen von Studenten des Lehramtes von Unterricht (White & Gunstone 1992).

Dieses Beispiel zu Vorstellungen von Studenten wird durch Befunde der TIMS-Studie bestätigt (Baumert et al. 1997; vgl. Gruber et al. 2000, S. 140). Für Deutschland und die USA wird Mathematikunterricht als „Wissenserwerbunterricht“ (Baumert et al. 1997, S. 215) bezeichnet. In Japan, USA und Deutschland bestimmt der Lehrer den Unterricht und mindestens 70% aller Äußerungen in einer Unterrichtsstunde entfallen auf ihn (Baumert et al. 1997, S. 231 f.; vgl. Gruber et al. 2000, S. 140). Eigene Erfahrungen als Lehrer an vier Gymnasien in Bayern von 1996 bis 2003 unterstreichen dieses Bild. Instruktionsbestimmte Lehr-Lern-Prozesse dominieren meist den Schulunterricht der Gegenwart in Deutschland.

2.2.2. Zu Vorläufern konstruktivistischer Positionen

Konstruktivistische Ansätze sind ebenfalls in den Kontext vorheriger z.T. historischer Vorstellungen zu stellen. Reinmann-Rothmeier & Mandl (2001, S. 621 ff.) nennen als Vorbilder z.B. Kerschensteiners „Arbeitsschule“, Bruners „Entdeckendes Lernen“, den „Amerikanischen Pragmatismus“ nach Dewey, die Projektmethode nach Frey, u.a.

Diese Auswahl an Vorläufern sei im Folgenden jeweils in Bezug auf die Lerntheorie des Konstruktivismus kurz vorgestellt.

Die Projektmethode nach Frey gilt nach Apel & Knoll (2001, S. 51) „bis heute [als] das Standardwerk zum Thema [Projektunterricht]“. Sie beschreibt informellen, weitgehend schülerbestimmten Unterricht mit dem Lehrenden in der Rolle eines „Hintergrundlehrers“ (Frey 1982). Schülerbestimmtheit beim Lernen mit einem Lehrer, der sich passiv bzw. „re-aktiv“ verhält, das sind Positionen wie sie die Lerntheorie des Konstruktivismus ebenfalls vertritt. Frey (1982, S. 46 und 208 f.) bezieht sich u.a. auf Dewey und sieht Verwandtschaft seiner „Projektmethode“ mit dem „Entdeckenden Lernen“.

Den Grundgedanken „Entdeckenden Lernens“ formuliert Neber (1973, S. 7) folgendermaßen: „Die Lernenden sollen ihr Wissen durch eigene Aktivität aufbauen, Fakten und Zusammenhänge selbständig suchen und ihre Lernvoraussetzungen zur Erweiterung ihrer Kenntnisse produktiv einsetzen (...)“. Die Nähe zum Konstruktivismus

wird noch deutlicher bei Bruner (1973, S. 20) selbst: „(...) ich meine, wenn man das Entdecken beim Lernen betont, so wirkt sich das auf den Lernenden gerade so aus, dass aus ihm ein Konstrukteur wird.“ Ohne sich auf die identische erkenntnistheoretische Ausgangsposition zu beziehen, verwenden beide Konzepte, Konstruktivismus und Entdeckendes Lernen, die gleiche Metapher: Konstruktion. Das Phänomen des „trägen Wissens“ (Gruber, Mandl & Renkl 2000, S. 139 ff.) ist ein Hauptkritikpunkt von Vertretern konstruktivistischer Lerntheorien an den Lernresultaten traditioneller Konzepte. Bruner (1973, S. 20) deutet dieses Phänomen bereits an und formuliert einen Ansatz seiner Überwindung: „Die Übung im Selbstentdecken lehrt (...), Information so zu erwerben, dass sie für das Problemlösen weitaus fruchtbarer wird.“ Die Verwendung des Komparatives „fruchtbarer“ deutet an, dass Bruner klar ist, dass „herkömmliche“ Wissensvermittlung Defizite bei den Lernern erzeugt, bezüglich der Kompetenz bestimmte Aufgabentypen zu lösen. Seine Lösung lautet „Selbstentdecken“. Reinmann-Rothmeier & Mandl (1994, S. 43) begegnen dem Problem des „trägen Wissens“ durch Authentizität und situierte Anwendungskontexte. Diese beiden Lösungsansätze zum Problem des „trägen Wissens“, Selbstentdecken bzw. Authentizität und Situietheit, schließen sich nicht aus. „Entdeckendes Lernen“ hat viele Positionen der Lerntheorie des Konstruktivismus vorweggenommen.

Lange vor Bruner pflichtet Kerschensteiner (1921, S. 90 f.) Deweys harter Kritik an der nur „aufs Hören“ eingerichteten Schule bei und fährt selbst fort: „Der Schüler (...) ist (...) nicht bloß aufs Hören und eine passive Aufnahme von fremdem Wissen eingerichtet.“ Das sind Gedanken, die noch heute Grundpositionen der Lerntheorie des Konstruktivismus umreißen. Kerschensteiner (1921, S. 94) fordert in der Schule „Arbeitsräume“ und „Lernräume“. Das ist konkret zu verstehen als unterschiedliche Zimmer, zwischen denen der Schüler hin und her wechseln kann, um Gelerntes auszuprobieren und neues Wissen selbst zu verlangen, das er dann sofort wieder erproben kann. Diese Elemente Kerschensteiners Vorstellungen, Selbstbestimmung, Eigenaktivität und Möglichkeiten zur Konstruktion eigener Wirklichkeiten, sind wichtige Merkmale der aktuellen Lerntheorie des Konstruktivismus.

Der „pädagogischer Bereich“ des Pragmatismus in den USA „verbindet sich (...) v.a. mit dem Name John Dewey“ (Frey 1982, S. 33). „Dewey schlägt schon in seiner Schrift ‚Demokratie und Erziehung‘ (1916) praxisorientierte Curriculumentwicklung

vor“ (Frey 1982, S. 34). Gefordert werden Einbindung „realer und sozialer Kontexte als Lernmedium“ und „Aufgaben mit Lebensnähe“ (Frey 1982, S. 33). Law (2000, S. 280) über Dewey: Wissen und Handeln „definieren sich gegenseitig durch die Erfahrung und den Kontext“. Folglich „muss nach Dewey“ der „Weitervermittlung unkritischer und unüberprüfbarer Wissens“ „entgegengewirkt werden“ (Frey 1982, S. 33). Diesen Elementen des Pragmatismus steht die Lerntheorie des Konstruktivismus sehr nahe.

2.2.3. Fazit

Instruktionsbestimmte Vorstellungen wie auch Lerntheorien des Konstruktivismus haben historische Vorläufer. Aktuell wird die Didaktik der Naturwissenschaften stark von konstruktivistischen Positionen geprägt. Der noch immer traditionell orientierte Schulunterricht folgt jedoch eher instruktionsbestimmten Vorstellungen von Lehr-Lern-Prozessen.

2.3. Aktuelle Lerntheorien des Konstruktivismus

2.3.1. Darstellung einer Auswahl aktueller Theorien

Die folgenden Ausführungen sollen inhaltlich den Kontext der verwendeten Theorie umreißen und gleichzeitig einen Eindruck von Dynamik und Vielfalt der aktuellen Vorstellungen zur Lerntheorie des Konstruktivismus geben. Die Theorie ist mit dem Begriff „Situating Cognition“ verknüpft.

Aus Mandl et al. (1997, S. 170) und Reinmann-Rothmeier & Mandl (2001, S. 615) werden gemeinsame Merkmale dieser Bewegung zusammengetragen:

- „Das Wissen in einer Gesellschaft stellt immer ‚geteiltes Wissen‘ dar“ (Reinmann-Rothmeier & Mandl 2001, S. 615; ähnlich Mandl et al. 1997).
- „Lernen ist stets situiert“ (Reinmann-Rothmeier & Mandl 2001, S. 615; ähnlich Mandl et al. 1997).
- „Wissen wird durch wahrnehmende Subjekte konstruiert“ (Mandl et al. 1997, S. 170; ähnlich Reinmann-Rothmeier & Mandl 2001).

Schon diese wenigen Punkte zeigen, Lernen nach der „Situating Cognition“-Bewegung heißt Lernen nach dem „Primat der Konstruktion“. Nicht alle genannten Merkmale der z.T. unterschiedlichen Autorenteams sind identisch. Mandl et al. (1997, S. 170) nennen noch: „Situierendes Wissen wird unter dem Anwendungsaspekt und damit unter dem Gesichtspunkt seiner Authentizität analysiert.“ Reinmann-Rothmeier & Mandl (2001, S. 615) dagegen: „Das konkrete Denken und Handeln eines Individuums lässt sich jeweils nur auf dem Hintergrund eines konkreten (sozialen) Kontextes verstehen.“ Zentrale Widersprüche der beiden Darstellungen sind damit nicht entdeckt. Im Gesamtbild unterscheidet sich die Setzung der Schwerpunkte in Nuancen.

Die Einordnung dieses Begriffes ist noch mehr im Fluss. Während Gerstenmaier & Mandl (1995, S. 870 f.) unter dem Oberbegriff „Neuer Konstruktivismus“ u.a. die Ansätze der „situierenden Kognition“ mit Clancey und Greeno und die der „Communities of Practice“ mit Lave, Rogoff und Suchman zusammenfassen, sprechen Mandl et al. (1997, S. 170) bei Lave, Rogoff und Greeno (außerdem bei Resnick) von „Ansätzen situierenden Lernens“. Reinmann-Rothmeier & Mandl (2001, S. 623) sehen dieselben vier Autoren als die Vertreter der „Situating Cognition-Bewegung“. Die Konzeptionen von Rogoff, Lave, Greeno und Resnick haben jeweils eigene Charakteristiken (Reinmann-Rothmeier & Mandl 2001, S. 615 f.).

Rogoff wird die Konzeption der ‚*Guided Participation*‘ zugeschrieben. Das „soziale Milieu“ schreibt „den Menschen eine Art kulturelles Curriculum vor“ (Mandl et al. 1997, S.169). Lernen erfolgt am besten durch von kompetenten Sozialpartnern unterstützte und „geleitete Teilnahme an sozialen Prozessen“ (Mandl et al. 1997, S.169). Sie selbst beschreibt ihren Ansatz wie folgt: „I use the analogy of apprenticeship to focus on how the development of skill involves active learners observing and participating in organized cultural activity with the guidance and challenge of other people“ (Rogoff 1990, S. 19).

„Der *Community of Practice*-Ansatz von Lave“ (Reinmann-Rothmeier & Mandl 2001, S. 615) beschreibt Lernen als dialektischen Prozess in alltäglicher Situation (Mandl et al. 1997, S.169).

Greenos Konzept der *Situiertheit* schildert die Beeinflussung kognitiver Prozesse durch Handlungseinschränkungen und Handlungsangebote der Situation (Reinmann-Rothmeier & Mandl 2001, S. 616). „Die Übertragung von Gelerntem auf neue Situationen kann gelingen, wenn entweder die erforderliche Aktivität eine Interaktion mit unveränderten situativen Handlungsangeboten beinhaltet oder wenn die situativen Handlungsangebote sich ändern und die Aktivität entsprechend transformiert werden kann“ (Mandl et al. 1997, S. 169). Vereinfacht: Sind die Handlungsangebote der neuen Situation identisch mit der, in der das Wissen erworben wurde, so kann Transfer gelingen. Sind die Handlungsangebote der neuen Situation verändert, so ist Transfer nur dann erfolgreich, wenn der Lerner flexibles Wissen zur Verfügung hat.

„Der ‚*Situated Cognition*‘ Ansatz von *Resnick*“ fasst Kognition als sozial geteilte Aktivität auf“ (Mandl et al. 1997, S. 170). Sie kritisiert das Lernen in der Schule scharf: „Im Gegensatz zum Lernen außerhalb der Schule arbeiten die Schüler weitgehend isoliert und werden auch isoliert beurteilt; Hilfsmittel dürfen nicht gebraucht werden; symbolisch vermittelte Inhalte stehen im Vordergrund und der Anwendungsbereich des Gelernten bleibt weitgehend unbeachtet“ (Reinmann-Rothmeier & Mandl 2001, S. 616).

„*Situated Cognition*“ ist ein Begriff, der vom „Unterpunkt“ zur „Hauptüberschrift“ wird, beides gleichzeitig sein kann und unter dessen erweiterter Bedeutung mal mehr mal weniger Vertreter subsumiert werden. Der Begriffsbildungsprozess, konstruktivistisch interpretiert, der Prozess eine intersubjektive Übereinkunft zu finden, ist anscheinend noch am Werk.

Zur „*Situated Cognition*“-Bewegung gehörig (Mandl et al. 1997) und gleichzeitig als „neue konstruktivistische Ansätze“ (Reinmann-Rothmeier & Mandl 2001) sind die Konzepte „*Cognitive Apprenticeship*“, „*Cognitive Flexibility*“ und „*Anchored Instruction*“ zu nennen.

Law (2000) stellt ebenfalls Vorläufer des Konstruktivismus und gegenwärtige Vertreter vor. Bis auf Dewey unterscheiden sich die historischen Vertreter von denen, die Reinmann-Rothmeier & Mandl (2001) nennen. Die aktuellen Vertreter der „*Situated Cognition*“ sind nur z.T. dieselben. Auch Law (2000, S. 257) thematisiert die Heterogenität der Begriffsdefinition von „*Situated Cognition*“.

Gunstone (2000), den Fokus eher auf „*Science Education*“ gerichtet, referiert ebenfalls konstruktivistische Strömungen. In keinem Punkt überlappen sich seine Referenzen mit denen Mandls, Gerstenmaiers, Reinmann-Rothmeiers etc. „Seine“ aktuellen konstruktivistischen Richtungen heißen „*The Leeds Programm* (...)“, „*The Waikato Group*“ und „*The Monash Group*“.

2.3.2. Kritik der dichotomen Positionierung der aktuellen Theorien nach den Etiketten „konstruktivistisch“ und „instruktionsbestimmt“ (bzw. traditionell, lehrerzentriert, sachorientiert, o.ä.)

Extreme Vertreter der Lerntheorie des Konstruktivismus gehen von einer weitgehenden Nicht-Vorhersagbarkeit der Wirkung von Instruktionen auf Lernende aus. Man

nimmt an, Lerner seien in ihrer Konstruktion völlig autonom. Planbarkeit jeglicher instruktionaler Maßnahmen ist damit nicht gegeben (Hoops 1998, S. 236). „Dem traditionellen Instructional Design wird generell rigorose externe Kontrolle unterstellt, von der sich die prinzipielle konstruktivistische Liberalität wohltuend abhebt“ (Hoops 1998, S. 239). Der Hintergrund dieser Ablehnung liegt in der Verengung der Theorien auf ein Entweder / Oder zweier Grundpositionen, nämlich traditionellen, zielorientierten Instruktionsmodellen und der Gegenposition, dem Konstruktivismus als didaktischer Leitlinie bei Lernprozessen (vgl. Hoops 1998, S. 247). Sicher liegt im Kern dieser dichotomen Positionierung, die durch die Konstruktivisten betrieben wird (Hoops 1998, S. 236), ein begründetes Anliegen.

Reine Instruktionsmodelle haben die Individualität und die Notwendigkeit der Berücksichtigung individueller Konstruktion vernachlässigt. Skinners (1971, S. 12) Rat bezüglich „weniger aktiver“ Schüler: „(...) der Lehrer [muss] ihm Fakten einprägen, ihm Vorstellungen beibringen, ihm gutem Geschmack oder Liebe zum Lernen einschärfen.“ Die Vorstellungen Skinners (1971) von Programmierem Unterricht, die einen Lerner im Zentrum eines optimal planbaren Prozesses sehen, im Extrem von Lernmaschinen „unterrichtet“, rufen bei Konstruktivisten großen Widerstand hervor.

Auf der anderen Seite muss man die oben angesprochene Ablehnung der Konstruktivisten über die Vorhersagbarkeit der Wirkung instruktionaler Maßnahmen ebenfalls sehr kritisch sehen. „[Es] lässt sich kaum bestreiten, dass Lehre häufig im Großen und Ganzen zu den geplanten Wirkungen führt (...)“ (Hoops 1998, S. 236). Wenn „die Auswirkungen von Instruktion (...) nicht vorhersehbar“ sind, so ist „Kommunikation (...) generell zum Scheitern verurteilt“ (Hoops 1998, S. 236). Die Vertreter des Konstruktivismus haben hier - offensichtlich mit erkenntnistheoretischer Schützenhilfe argumentierend, denn daher leitet sich der Gedanke der Autonomie des Lerners ab (Hoops 1998, S. 238) - die Prämisse der Notwendigkeit individueller Konstruktion überbeansprucht. Beide Positionen, Konstruktivismus und „Instruktivismus“, stellen in ihren extremen Ausprägungen anscheinend keine sehr brauchbare konzeptuelle Basis für die praktische Verbesserung von Lernprozessen dar.

Bei näherer Betrachtung finden sich auf der Ebene der historischen Vorläufer sehr deutlich Zwischentöne. Herausgegriffen seien Skinner (Programmierter Unterricht), ein Vertreter, der klar Reinmann-Rothmeiers & Mandls (2001) „Primat der Instruktion“ zuzuordnen ist, und Bruner (Entdeckendes Lernen), dem „Primat der Konstruktion“ zugerechnet.

Bruner (1967) widerspricht schon in seiner Beschreibung der Rolle von Lehrer und Schüler der erkenntnistheoretischen Sicht des Konstruktivismus über das Wissen. Den Lehrer beschreibt er als „one who possesses something“, den Schüler als „one who does not“ (Bruner 1967, S. 42). Wissen, in dieser deutlichen Form als Besitz gekennzeichnet, entspricht nicht konstruktivistischen Sichtweisen, sondern eher dem Objektivismus. Einschränkend ist zu sagen, dass Bruner selbst sich nicht auf den Konstruktivismus beruft. In der Konkretisierung seiner Ideen verlässt Bruner ebenfalls puristische konstruktivistische Vorstellungen: „Since learning and problem solving depend upon the exploration of alternatives, instructions must facilitate and regulate the exploration of alternatives on the part of the learner“ (Bruner 1967, S. 43). Ein gewisses Maß an Instruktion hält Bruner offensichtlich für erforderlich. Bruner (1967, S. 49) spezifiziert: „Instruction consists of leading the learner through a sequence of statements (...) of a (...) body of knowledge (...)“. Wissen als Besitz, Instruktion als geforderte Maßnahme, bestimmte, vom Lehrer für richtig gehaltene Prozesse, herbeizuführen. Sicher ist Bruner ein wichtiger Vorläufer der Lerntheorie des Konstruktivismus, folgt selbst aber nicht rein konstruktivistischen Vorstellungen.

Inwieweit erweist sich Skinners Standpunkt als „anti-konstruktivistisch“? „Wenn wir für Lehren eine Metapher suchen, so bietet sich das Wort Instruktion, Unterweisung – oder besser Konstruktion, Aufbau – an“ (Skinner 1971, S. 13). – Seltsame Worte für den „typischsten“ Vertreter der Gegentheorie zu der Lerntheorie des Konstruktivismus. Weiter liest man unter der Überschrift „Wir lernen durch Tun“: „Es ist wichtig zu betonen, dass ein Schüler Wissen aus einer Umwelt nicht passiv absorbiert, sondern eine aktive Rolle dabei spielen muss (...)“ (Skinner 1971, S. 14). Dies ist eine auch von Konstruktivisten vertretene Meinung. „Der Schüler muss die Welt, in der er lebt, kennen lernen und muss deshalb mit ihr in Berührung gebracht werden“ (Skinner 1971, S. 15). Wichtig ist nach Skinner (1971, S.13) eine „Gelegenheit, in der sich ein Verhalten ereignet.“ – Konstruktivisten bezeichnen Ähnliches als Authentizität und Situiertheit. Mit einem weiteren Zitat Skinners (1971, S. 17) sollen diese Ausführungen beendet werden: „Schon lange wird darauf hingewiesen, dass ein Lebewesen hauptsächlich dadurch lernt, dass man seine Umwelt verändert.“ Aus Sicht des Konstruktivismus interpretiert, beschreibt Skinner „Perturbation“ (vgl. Maturana & Varela 1987). Das vorherrschende Bild der Wirklichkeit muss, wenn man aus konstruktivistischer Sichtweise Veränderung intendiert, durch Konfrontation mit Phänomenen, die nicht in dieses Bild passen, erschüttert werden. Diese Perturbation führt zu entsprechender Korrektur der subjektiven Wirklichkeit. Sicher war Skinner trotz dieser Beispiele kein Konstruktivist. Die aufgeführten Positionen sollen dazu beitragen, die Einseitigkeit der Perspektive aufzuweichen.

Die Dichotomie zwischen den Theorien eines „Primats der Instruktion“ bzw. „der Konstruktion“ sind in der dargestellten Pointierung nicht aufrecht zu erhalten. „Konstruktion und Instruktion lassen sich nicht nach einem Alles-oder-Nichts-Prinzip realisieren“ (Reinmann-Rothmeier & Mandl 2001, S. 627). Skinner wie Bruner hatten sicher Erfahrung mit realen Lernsituationen und vielleicht bewirkt dieses „Eichen“ an der Realität die geschilderte Aufweichung der Extrempositionen. Die klare Trennung zwischen Instruktion und Konstruktion, die beschriebene Dichotomie sollte man vielleicht ersetzen durch ein „bipolares Kontinuum“ (vgl. Hoops 1998, S. 247), dessen Endpunkte „Primat der Instruktion“ bzw. „der Konstruktion“ heißen können. Die meisten umgesetzten Vorstellungen zu Lernprozessen dürften zwischen diesen Endpunkten anzusiedeln sein.

Zwei Beispiele folgen: Vance (1995) stellt zwei in der Schule erprobte Methoden, die weitgehend Prinzipien der Lerntheorien des Konstruktivismus entsprechen, vor: „the newspaper article and [the] imaginative story approach“. Trotz Selbstbestimmtheit, Selbsttätigkeit und der Möglichkeit für die Schüler, eigene Wirklichkeiten aufzubauen, fehlten nicht selten entscheidende Informationen oder Impulse, die ökologischen Prinzipien, um die es hier ging, fachlich richtig zu entwickeln. „Sanfte“ instruktionale Hilfen waren unumgänglich. Lord (1997) realisiert im College zwei Methoden für Biologieunterricht: „the traditional teacher-centered manner of lecture“ und „the student-centered, constructivist format“. Selbst sein konstruktivistischer Ansatz ist geprägt durch ein instruktionales Korsett, das die erforderliche inhaltliche Planbarkeit der Unterrichtsstunden gewährleistet. Diese Beispiele sind typisch für Umsetzungen konstruktivistischer Ideen im Unterricht. In den hier rezipierten Untersuchungen werden keine Reinformen, die puristischen konstruktivistischen Vorstellungen folgen, in Schule oder Universität umgesetzt.

2.3.3. Konstruktivismus in Didaktik und Unterricht, ein Dilemma

Jede Kommunikation beruht auf „intersubjektiver Übereinkunft“ (Hansmann 2002, mündliche Mitteilung), auf sozialer Konstruktion. Driver (1995, S. 392 ff.) beschreibt dies als „interpersonal construction of knowledge“. Es geht um Konstrukte, die von mehreren Subjekten in wichtigen Punkten gemeinsam getragen werden. Viele Subjekte können z.B. darin übereinkommen, dass bestimmte Wissensinhalte, welche selbst Konstruktionen darstellen und selbst schon Resultate intersubjektiver Übereinkünfte sind, für das aktuelle und zukünftige Leben von Menschen bedeutsam sind. „Welches Wissen ist notwendig, um so handeln zu können, dass das Leben gelingt?“ Das ist eine der zentralen Fragen der Didaktik (Ladenthin 2001, S. 306). Diese Frage nach den wichtigen Inhalten stellt sich in jedem Fach, natürlich auch in der Biologie.

Die zentralen Inhalte möglichst effektiv und effizient zu vermitteln, beschreibt die zweite zentrale Frage der Didaktik: „Wie macht man Erkenntnis zugänglich?“ (Ladenthin 2001, S. 307). Vielleicht etwas schlichter formuliert: Es geht um die Frage nach dem Wie des Lernens. Nach konstruktivistischer Denkart steht man nun vor einem Dilemma: Man möchte von einem Wissenden zu einem Unwissenden planvoll etwas weitergeben, das aber im Wissen um die Notwendigkeit individueller Konstruktionsprozesse. Einerseits sollen die Lerntheorien des Konstruktivismus unterrichtlich Anwendung finden, können sich damit einer Tradition des Unterrichts nicht verschließen, und nicht der Grundidee einer planvollen Vermittlung von Inhalten. Andererseits leiten sich Lerntheorien des Konstruktivismus aus der entsprechenden Erkenntnistheorie ab (Hoops 1998, S. 246) und sind damit Prinzipien autonomer Konstruktion verpflichtet. Die folgende Aussage pointiert die „Zwickmühle“: Entweder ein Schüler

lernt gut (nämlich nach konstruktivistischen Vorstellungen) oder man bestimmt, was er lernen soll (intentionale Weitergabe von Wissen). Betrachtet man dieses Dilemma, so könnte man salopp fragen: Wo liegt der Fehler? Sollte man lieber nicht planvoll Wissen vermitteln, oder stimmt vielleicht mit der „abgeleiteten“ Lerntheorie etwas nicht? Für Hoops (1998, S. 246 f.) ist die Entscheidung eindeutig: Er kritisiert das „Deduktionspostulat“, die Ableitung der Lerntheorie aus der Erkenntnistheorie des Konstruktivismus.

2.3.4. Kritik der Lerntheorie des Konstruktivismus

Konstruktion im Sinne der Erkenntnistheorie des Konstruktivismus lässt sich anschaulich mit Hilfe von Kommunikationsprozessen (Lernprozesse zählen dazu!) erläutern. Man kann zeigen, wie Kommunikation gelingt (Beispiel: Lehrerfortbildung Hansmann 2002) oder wie Kommunikation misslingt (Beispiel: Fabel⁵ von Lionni (1970) im Vortrag von Mandl zum 19. Lehrertag Oberfranken, veranstaltet vom Bayerischen Lehrer- und Lehrerinnenverband e.V. im Jahre 2001). – Beides lässt sich nutzen, das Zustandekommen subjektiver Konstruktion abzuleiten. Beide Male bedient man sich der Grundidee einer Parabel: Ein leicht nachvollziehbarer Kommunikationsprozess wird geschildert. Zentrale Elemente dieses Prozesses werden auf eine andere Ebene bezogen, gewissermaßen gespiegelt, so dass die Prinzipien auf dieser neuen Ebene verstehbar werden. - Kommunikation und Lernen werden benutzt, um erkenntnistheoretische konstruktivistische Kerninhalte, nämlich Existenz und Zustandekommen subjektiver Konstruktionen, besser zugänglich zu machen.

Es ist verlockend, die Parabel gewissermaßen ein weiteres Mal zu spiegeln und aus erkenntnistheoretischen Sätzen, ideale Bedingungen für Lernsituationen abzuleiten. Diese Deduktion führt jedoch in die Irre.

⁵ Frosch und Fisch unterhalten sich: Die Landlebewesen, von denen der Frosch erzählt, z.B. von einer Kuh, stellt sich jedoch der Fisch jeweils als Fische mit zusätzlichen Eigenschaften vor, z.B. ein Fisch mit Beinen, Hörnern und Euter. – Die Kommunikation scheitert an den Beschränkungen der Wirklichkeit des Fisches.

Sicher kann man sich vorstellen, ideale Bedingungen für Lernprozesse zu schaffen, in denen subjektive Konstruktion besonders gut gelingen kann. Im Umkehrschluss, Methoden, die von Lerntheorien des Konstruktivismus abweichen, das Vorhandensein subjektiver Konstruktion abzusprechen, widerspricht der Grundaussage der Erkenntnistheorie des Konstruktivismus. Erkenntnistheoretischer Konstruktivismus hat nicht selektiv für bestimmte Arten von Lernprozessen oder durch bestimmte Methoden herbeigeführtes Lernen oder in bestimmten Lernumgebungen stattfindendem Lernen Gültigkeit, sondern in jedem Fall (vgl. Hoops 1998, S. 238). Einen Beleg dafür liefert die Aussage des Neuropsychologen Spitzer (2002, S. 11): „Unser Gehirn lernt immer.“ Es ist für ein menschliches Gehirn nicht möglich, nicht zu lernen (vgl. Spitzer 2004, S. XVI). Damit ist es nicht möglich, nicht zu konstruieren (im Sinne von: subjektive Konstruktionen aufzubauen). Wenn jedoch unterrichtliche Situationen, die explizit nicht konstruktivistischen Positionen folgen, gleichwohl zu „erfolgreichen“ subjektiven Konstruktionen führen, wenn gelernt wird, worin unterscheiden sie sich dann von unterrichtlichen Situationen, die in idealer Weise konstruktivistischen Lerntheorien entsprechen? Es gibt aus der Warte der Erkenntnistheorie des Konstruktivismus keinen qualitativen Unterschied (vgl. Hoops 1998, S. 237)! Es ist nicht theoretisch begründbar, „konstruktivistisches Lernen“ von Lernen nach anderen Lerntheorien abzugrenzen. Darum wird dieser Begriff „konstruktivistisches Lernen“ vermieden, es sei denn als Teil von Zitaten. Dieser Terminus stellt eine Tautologie dar. Lernen bedeutet, Informationen in seine subjektive Konstruktion zu übernehmen, nicht-konstruktivistisches Lernen kann es demnach – im Sinne der Erkenntnistheorie des Konstruktivismus - nicht geben (vgl. Hoops 1998, S. 238).

2.3.5. Die gemäßigt konstruktivistische Position

Reinmann-Rothmeier & Mandl (2001) verlassen ebenfalls den Pfad der dichotomen Positionierung: Instruktion *oder* Konstruktion. „(...) der Vorgang des Instruierens seitens der Lehrenden und der Vorgang des Konstruierens seitens des Lernenden [schließen] einander [nicht aus]“ (Reinmann-Rothmeier & Mandl 2001, S. 625). White & Gunstone nennen bereits 1992 (S. 13) als einen von „three ways in which construction of meaning occurs“, nämlich „under the guidance of a teacher.“ Reinmann-Rothmeier & Mandl (2001, S. 626) schlagen statt eines radikalen Standpunktes gegen den „Instruktivismus“ eine „pragmatische“, „gemäßigt konstruktivistische Positi-

on“ vor und folgen damit nicht Hoops (1998, S. 249) Vorschlag auf den Terminus „Konstruktivismus“ ganz zu verzichten. Ob und inwieweit der erkenntnistheoretische Hintergrund (mit all seinen Fallstricken, vgl. oben) bei diesem Vorschlag noch eine Rolle spielt, soll nicht weiter diskutiert werden. Entscheidend ist vielmehr die konkrete Ausgestaltung dieser Lernform und deren empirische Überprüfung (vgl. Abbildung 7). Es geht um „gezielte externe Einflüsse auf die Lernenden und wissenschaftlich zu prüfende Hypothesen darüber, wie dies unter unterschiedlichen Bedingungen am besten geschehen sollte“ (Hoops 1998, S. 249), eine Rückbesinnung auf konkrete Forschungshypothesen, -designs und entsprechende Befunde. Wie sehen nun die Merkmale dieser pragmatischen Position aus?

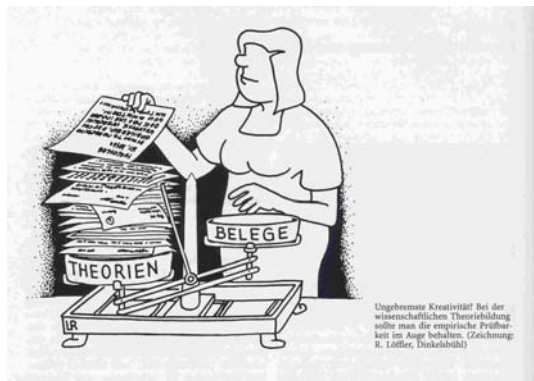


Abbildung 7: Situation in der Lehr-Lern-Forschung (Witzzeichnung von Löffler, Dinkelsbühl in Bortz & Döring, 1995, S. 328).

Kriterien der gemäßigten Theorie: Folgende „Prozessmerkmale“ gelten für Lernprozesse (verändert nach Reinmann-Rothmeier & Mandl 2001, S. 626; 1994, S. 44 ff.):

- Selbststeuerung
- Eigenaktivität
- Situiertheit (unter Einbeziehung der sozialen Dimension)
- Konstruktion eigener Wirklichkeit

Zur Selbststeuerung:

„Der Begriff des selbstgesteuerten Lernens kann gegenüber verwandten Termini wie autonomem, selbstbestimmtem, selbst organisiertem oder autodidaktischem Lernen nicht scharf abgegrenzt werden“ (Reinmann-Rothmeier & Mandl 2001, S. 632). Selbststeuerung ist ein Thema unterschiedlicher psychologischer Blickwinkel (Reinmann-Rothmeier & Mandl 2001, S. 632), neben Reinmann-Rothmeier & Mandl z.B.

auch von Deci & Ryan (2000 und 1993) und Spitzer (2004) bearbeitet. In dieser Arbeit sei darunter Lernen verstanden, das „durch einen hohen Grad an Autonomie“ des Lerners ausgezeichnet ist, selbst wenn „einzelne Handlungsabschnitte fremdbestimmt sind“ (Reinmann-Rothmeier & Mandl 2001, S. 632). „Voraussetzung (...) ist, dass der Lernende Wahlmöglichkeiten hat und ihm die Möglichkeit gegeben wird, selbst etwas zu bewirken“ (Reinmann-Rothmeier & Mandl 2001, S. 633). Folgende Handlungsabschnitte können selbstbestimmt sein (vgl. Reinmann-Rothmeier & Mandl 2001, S. 633): Der Lernende kann Einfluss auf die Zielsetzung nehmen, wodurch die intrinsische Motivation des Lerners positiv beeinflusst werden kann. Ähnliches äußert Cooper (1993) bezüglich Lernens nach konstruktivistischen Maximen über die Motivation der Lerner. Der eigentliche Prozess der aktiven Wissensaufnahme bzw. Informationsverarbeitung kann nur durch den Lerner gesteuert sein. Das „Koordinieren des Lernens“, interne und externe Einbindung in die für das Lernvorhaben irrelevanten Umstände, kann selbstgesteuert sein. Ferner ist die Organisation des Lernprozesses ein möglicherweise selbstgesteuerter Vorgang. Die Überwachung der Erreichung des vielleicht selbstbestimmten Ziels und jede weitere Metaebene des Lernprozesses kann dem Lerner bekannt sein, wodurch seine subjektive und objektive Selbstbestimmung beeinflusst werden kann.

Zur Eigenaktivität:

Auf der Ebene der eigentlichen Wissensaufnahme bzw. Informationsverarbeitung kann nur Eigenaktivität zu erfolgreichem Lernen führen. Das erkannte bereits Skinner (1971, S. 14). Eigenaktivität kann jedoch auf anderer Ebene besondere Involviertheit bedeuten, z.B. psychomotorisches Engagement oder situatives Interesse.

Zur Situiertheit (unter Einbeziehung der sozialen Dimension):

Lernen findet immer in einem bestimmten Kontext statt, so dass es „an die inhaltlichen und sozialen Erfahrungen der Lernsituation gebunden ist“ (Reinmann-Rothmeier & Mandl 2001, S. 615). Damit kann man versuchen, die materiellen Bedingungen der Lernsituation für den Lernprozess möglichst günstig zu arrangieren, beispielsweise Situationen zu schaffen, die einem möglichen Anwendungskontext entsprechen, z.B. authentische Situationen, wie von Reinmann-Rothmeier & Mandl (1994, S. 46) gefordert. Die soziale Ebene umfasst die aktuelle soziale Situation zwischen Lehrer und Lerner und zwischen den Lernern, sowie den soziokulturellen Hintergrund der Lerner.

Zur Konstruktion eigener Wirklichkeiten:

Nach Überlegungen von Reinmann-Rothmeier & Mandl (2001) konstruieren Lerner unter bestimmten Bedingungen – genannt seien Situiertheit, Eigenaktivität, Selbststeuerung, etc. - leichter subjektive Wirklichkeiten, die zudem „besser geeignet“ sind, den Subjekten zu nützen, denn dieses Wissen ist gekennzeichnet durch Flexibilität in multiplen Anwendungskontexten. Das beschreibt der Terminus „nicht-träges Wissen“ (vgl. Reinmann-Rothmeier & Mandl 1994, S. 43; Gruber et al. 2000, S. 139). Dieses Prozessmerkmal, Konstruktion eigener Wirklichkeiten, ist jedoch das Ziel des Lernprozesses, und damit nicht geeignet den Prozess zu charakterisieren. Wie oben ausgeführt (vgl. Einleitung 2.3.4.), sind nach Grundsätzen der Erkenntnistheorie des Konstruktivismus Lernprozesse nach konstruktivistischen Maximen nicht theoretisch begründbar von anderen Lernprozessen abzugrenzen. Auch darum ist dieses Prozessmerkmal problematisch. Von diesen Schwächen abgesehen, sind die von Reinmann-Rothmeier & Mandl (2001, S. 626 ff.) aufgestellten Prozessmerkmale einer gemäßigt konstruktivistischen Lerntheorie nachvollziehbar, eindeutig und in sich konsistent.

II. Unterrichtliche Umsetzung

Es folgt die fachdidaktische Operationalisierung der in der Einleitung entwickelten Gedanken.

1. Voraussetzungen

Ziel der pragmatischen Position von Reinmann-Rothmeier & Mandl (2001, S. 645) ist es, „eine Balance zwischen Instruktion und Konstruktion zu erreichen.“ Anzustreben ist Lernen als Prozess nach Prinzipien des gemäßigten Konstruktivismus beschrieben durch die o. g. „Prozessmerkmale“ (Reinmann-Rothmeier & Mandl (2001, S. 626). Besonders entscheidenden Einfluss auf die Balance zwischen Instruktion und Konstruktion hat das Prozessmerkmal „Selbststeuerung“ (vgl. Killermann 2000, S. 24 f.; vgl. Häußler et al. 1998, S. 158). Vereinfacht formuliert, verändert ein höheres Maß an Selbstbestimmung den Lernprozess in Richtung konstruktivistischer Vorstellungen, ein höherer Grad an Fremdbestimmung in Richtung instruktionsbestimmter Lehr-Lern-Situationen.

In dieser Untersuchung wird dieses zentrale Prozessmerkmal, die Selbststeuerung der Lerner, herausgegriffen und variiert. Es soll im Rahmen einer Schulveranstaltung von der Dauer eines Vormittags⁶ der Unterrichtserfolg bezüglich des relativen Grades der Selbststeuerung von „sehr gering“ über „ein mittleres Maß“ bis „sehr hoch“ an je einer Schülerpopulation untersucht werden. Im Unterricht in einem außerschulischen Lernort, dem Naturkundemuseum des Umweltschutz-Informationszentrums Lindenhof in Bayreuth, werden dementsprechend drei Treatments umgesetzt.

1.1. Fachdidaktische Charakterisierung

Zunächst ist der Unterricht vorzubereiten. „Mit Unterricht sind (...) solche Situationen gemeint, in denen mit pädagogischer Absicht und in organisierter Weise innerhalb eines bestimmten institutionellen Rahmens von professionell tätigen Lehrenden Lernprozesse initiiert, gefördert und erleichtert werden“ (Reinmann-Rothmeier &

⁶ Es findet eine einmalige Intervention statt, da die inhaltliche Substanz des Museums nicht ausreicht, um Schulleiter, Lehrer, Eltern und Schüler für mehrmalige Museumsbesuche zu gewinnen.

Mandl 2001, S. 603). Die genannten Bedingungen sind erfüllt, damit kann berechtigter Weise von „Unterricht“ gesprochen werden.

Der Unterricht im Museum, im Folgenden auch als „Museumsbesuch“ bezeichnet, folgt – nicht zuletzt aus organisatorischen Gründen – nicht dem „üblichen Verlauf“ lehrergeleiteten Unterrichts. Die Grobeinteilung, die für praktisch jeden Unterricht gelten soll und die Eschenhagen et al. (2001, S. 170) mit „Einstieg“, „Be- oder Erarbeitung“ und „Ergebnissicherung und Vertiefung“ beschreiben, wird – betrachtet man den Museumsbesuch als Ganzes – durchbrochen. Sie findet sich aber durchaus bei der Bearbeitung der einzelnen Stationen wieder. Mit „Station“ ist jeweils ein Schaufenster (= Diorama oder Schaukästen), mit entsprechendem Aufgabenpaket gemeint.

Für das Unterrichten am außerschulischen Lernort ist – genauso wie für den Unterricht an der Schule im Klassenraum – eine Unterrichtsvorbereitung⁷ nötig. Man versteht darunter, „einen Plan im Sinne einer Gedankenskizze zu entwerfen“ (Eschenhagen et al. 2001, S. 153) und gleichzeitig die technischen und organisatorischen Grundlagen für den Unterricht bereit zu stellen. Genau genommen wird damit auch die mentale Vorbereitung des Lehrers und seine Vorbereitungen, eine „gute Atmosphäre in einer Lerngruppe“ zu schaffen, bezeichnet (Eschenhagen et al. 2001, S.153).

Besondere Umstände ergeben sich aus der Natur der Untersuchung: Durch die Treatments sind per se unterschiedliche Grade an Schülerzentrierung, im Sinne von Beteiligung der Schüler an der Bestimmung des Lerninhalts (Berck & Graf, 2003, S. 82), nämlich von gar keiner bis zu einem hohen Grad an Beteiligung, gegeben. Neben didaktischen Erwägungen sind methodische Überlegungen die Studie betreffend zu berücksichtigen. Beispielsweise ist es für die Vergleichbarkeit der Treatmentgruppen unabdingbar, so weit wie möglich Störfaktoren auszuschließen und äußere Bedingungen konstant zu halten, so v. a. die Bearbeitungszeiten der drei Treatments. Dies gibt dem Unterricht ein zeitliches Korsett, das flexibles Reagieren auf Schülerhandeln erschwert.

Der Unterricht besteht aus Sequenzen von je zehn Minuten. In dieser Zeit bearbeitet jeder der als Probanden gewählten Gymnasiasten der fünften Jahrgangsstufe jeweils in einer Kleingruppe von drei oder vier Kindern Arbeitsblätter, die unterschiedliche Grade an Selbstbestimmung zulassen. Ein traditioneller Lehrer, der in Frontalunterricht Informationen weitergibt, tritt nicht in Erscheinung. Eine Gruppe bestehend aus zwei oder drei Kleingruppen ist einer Betreuerin zugeordnet, die den organisatori-

⁷ Der Begriff Unterrichtsplanung wird nicht verwendet, da Eschenhagen et al. (2001, S. 153) darunter ausschließlich die „rein gedankliche Vorwegnahme von Unterricht“ verstehen, hier aber ein weiter verstandener Begriff von vorbereitenden Handlungen für den Unterricht gemeint ist.

schen Ablauf des Museumsbesuchs regelt. Die eigentliche Bearbeitung der Stationen bleibt jedoch ganz den Schülern überlassen.

Für die weitere fachdidaktische Charakterisierung von Unterricht kann man eine didaktische Analyse durchführen.

Darunter versteht man eine „eingehende Auseinandersetzung mit Zielen und Inhalten des Unterrichts sowie mit ihrer Begründung im Hinblick auf den Bildungsgehalt“ (Killermann 1995, S. 253; vgl. Staeck 1995, S. 295 ff.). Nach Klafki (1964, S. 14) besteht die Funktion einer didaktischen Analyse v. a. darin, den „allgemeinen Bildungsgehalt“ eines „besonderen Bildungsinhaltes“ zu ermitteln. Der Bildungsgehalt eines Inhalts setzt sich zusammen aus dem Bildungswert der Sache an sich, daraus, inwieweit der Inhalt im gegebenen gesellschaftlichen und kulturellen Umfeld als reales oder gewünschtes Bildungsgut, also geteiltes geistiges Eigentum, für den Lerner zu bewerten ist, und daraus, inwieweit der Inhalt wahrscheinlich Bezug zur Zukunft des Lerners hat (Klafki, 1964, S. 10 ff.).

Klafki (1964, S. 15 ff.) schlägt fünf Fragen vor, um den Bildungsgehalt einer Sache zu ergründen, die im Folgenden sinngemäß wiedergegeben werden:

- Welchen allgemeinen Sinnzusammenhang vertritt dieses Thema?
- Welche Bedeutung hat der Inhalt für das Leben der Lerner bzw. sollte er haben?
- Welche zukünftige Bedeutung hat der Inhalt wahrscheinlich für das Leben der Lerner? (Diese Frage wird gestellt in Hinblick auf das Ziel des gebildeten Laien.)
- Welches ist die (fachliche) Struktur des Inhalts?
- Durch welche Fälle/ Beispiele lässt sich der Inhalt dem Lerner begreiflich machen?

Ähnliches umschreiben Relevanzkriterien (Berck & Graf 1987, S. 9 ff.) bzw. Curriculumdeterminanten (Berck 2003) mit Fach-, Schüler- und Gesellschaftsrelevanz. Die Grundprinzipien der Stoffauswahl sind seit Klafki weitgehend konstant geblieben.

Auf eine vollständige didaktische Analyse soll verzichtet werden. Dies würde für die sechs Stationen sehr viel textlichen Raum in Anspruch nehmen. Es wird betont, dass bei der Konzeption des relativ neuen Naturkundemuseums des Umweltschutz- Informationszentrums Lindenhof (Konzipierung 1989 bis 1994), die in enger Abstimmung mit dem Lehrstuhl Didaktik der Biologie der Universität erfolgte, entsprechende Fragen z.B. bezüglich einer Beachtung der Curriculumdeterminanten hinreichend bedacht wurden. Zielformulierungen bezüglich der Treatments sind unter Kapitel 2.3.2. konzipiert. Wichtig erscheint dennoch eine fachliche Charakterisierung der In-

halte der ausgewählten Stationen. Diese so genannte Sachanalyse wird erweitert um eine besondere Berücksichtigung des musealen Kontextes und der Intention der Schülerexkursion.

1.2. Lernen von „Ökologie“ im Museum

1.2.1. Überlegungen zum Museumsbesuch einer Schulklasse

Bei einem Museumsbesuch lassen sich unterschiedliche Ebenen der Beeinflussung eines intendierten Lernprozesses von Besuchern denken. Man könnte die Ausstellung selbst verändern bzw. eine eigene Ausstellung nach bestimmten Kriterien entwerfen (wie z.B. Schmitt-Scheersoi & Vogt, 2003, S. 175 ff.). Dieser Aufwand ist i.d.R. von einem Lehrer für den Besuch eines außerschulischen Lernortes wie eines Museums nicht zu leisten. Darum wird die Ausstellung des Museums als gegeben angesehen, gestaltet wird lediglich die Art der Besuches. Dieser Ansatz entspricht den Möglichkeiten, mit Hilfe derer ein Lehrer einen Museumsbesuch in seiner täglichen Praxis umsetzen kann. Wie die in dieser Untersuchung gewählten Gestaltungen im Einzelnen aussehen, sei im folgenden Kapitel „Unterrichtliche Umsetzung“ ausgeführt. Zunächst sind die musealen Gegebenheiten zu erläutern.

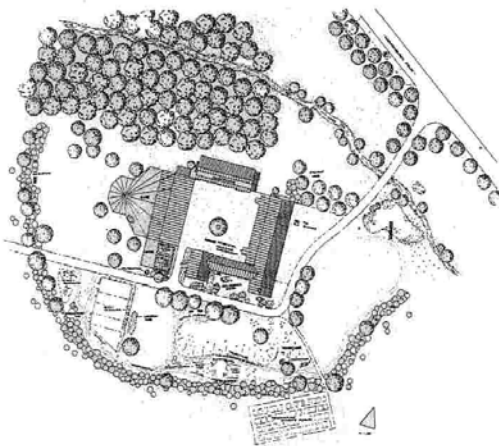
1.2.2. Museum

Vorab ist zu klären, was in dieser Untersuchung unter einem Museum verstanden wird. Ausgangspunkt hierfür ist die Definition des Deutschen Museumsbundes (1978, zitiert nach Frank 2001, S. 23 f.), die hier in verkürzter Form wiedergegeben wird: „Ein Museum ist eine (...) Sammlung“ „erhaltenswerter Objekte“ mit „fachbezogener Konzeption“ ohne „kommerzielle Funktion“, die „der Öffentlichkeit zugänglich ist“. Die „Objektsammlung des Museums muss fachmännisch betreut werden und wissenschaftlich ausgewertet werden können. Die Schausammlung des Museums muss eine eindeutige Bildungsfunktion besitzen“.

Das Umweltschutz-Informationszentrum Lindenhof ist ein ehemaliger Aussiedlerbauernhof im Süden Bayreuths (vgl. Abbildungen 8 und 9). Ein Naturerlebnisbereich von 17 ha bietet vielfältige Biotoptypen, z.B. Hecke, Wald, Trockenmauer, Teich, Fließ-

gewässer, und weitere Möglichkeiten für Kinder in „Unterricht“ oder Freizeit die Natur im Wirklichkeitszusammenhang (vgl. Uhlig 1962, S. 68) zu erfahren, z.B. einen Naturlernpfad und eine Wildbienenwand. Der „Lindenhof“ wird vom Landesbund für Vogelschutz (LBV) betrieben und wurde von der „Deutschen Bundesstiftung Umwelt“ maßgeblich unterstützt. In den „Leitlinien für die Förderung“ wird explizit die Vermittlung von „Wissen über die Umwelt“ verlangt. Zwischen LBV und Universität (Verantwortliche Koordination: Lehrstuhl für Didaktik der Biologie) gibt es einen Kooperationsvertrag, in dem die didaktisch-pädagogische Begleitung, incl. empirischer Unterrichtsforschungen, festgelegt sind.

Das Museum des Umweltschutz-Informationszentrums Lindenhof besteht aus zwei Teilen, die „umgangssprachlich“ als „Kinder- und Naturkundemuseum“ bezeichnet werden. Im Folgenden verwendete Begrifflichkeiten und geschilderte Konzepte beziehen sich z.T. auf persönliche Mitteilungen von Robert Pfeifer, der das Museum des Lindenhofs maßgeblich (mit-) konzipierte und aufbaute, und unveröffentlichten Unterlagen des Lehrstuhls der Didaktik der Biologie der Universität Bayreuth, z.B. obige Förderleitlinien, Kooperationsvertrag und Unterlagen aus der Planungsphase zum „Projekt Lindenhof“ (Didaktik der Biologie: unveröffentlichte Manuskripte ab 1989).



Abbildungen 8 und 9: Das Umweltschutz-Informationszentrum Lindenhof: links eine Übersichtsskizze, rechts die Außenansicht (aus den oben genannten Unterlagen).

Das „Kindermuseum“ ist ein hoher Raum aus dem Dachstuhl einer Scheune (vgl. Abbildung 10 bis 17). Durch einen Zwischenboden ist ein Teil des Giebels abgeteilt

und begehbar („Baumhaus“). Entlang der Wände und unter den Schrägen des freien Teils des Giebels sind ca. 30 Stopfpräparate heimischer Vögel und Säugetiere ausgestellt. An Strukturelementen gibt es einen nachgebauten Dachsbau mit einer „bekrabbelbaren“ Höhle und eine als Kletterbaum gestaltete „Leiter“. In der Höhle ist nach der ersten Biegung - hinter Glas und beleuchtet - ein Dachs ausgestellt. Der Kletterbaum führt zu dem begehbaren Giebel mit schrägen Wänden und einer Grundfläche von etwa zehn Quadratmetern („Baumhaus“). Des Weiteren gibt es im Kindermuseum verschiedene Sitzgelegenheiten aus Holz, kleine Hocker mit Tiermotiven und Bänkchen, z.B. gestaltet als Raupe. Keines der ausgestellten Stopfpräparate ist mit Trivialnamen oder binärer Nomenklatur nach Linne` bezeichnet. Die einzige Möglichkeit für den Museumsbesucher einige der Trivialnamen zu erfahren, besteht darin, einen elektronischen Lernapparat (Ecoboard) zu nutzen, wobei richtig benannte Stopfpräparate durch ein Lichtsignal angestrahlt werden. Als weiteres Lernspiel gibt es im Kindermuseum Tierstimmen-Telefone. Das sind „herkömmliche Telefone“, bei denen durch das Wählen bestimmter zweistelliger Nummern die Stimmen von Vögeln, Amphibien und einer Walart (Buckelwal) zu hören sind.

Sich an den Ausführungen von König, G. (1999, v. a. S. 159) orientierend, z.B. dass Kindermuseen „besucherorientierte Kulturorte“ seien, ist die für diesen Ausstellungsteil gewählte Bezeichnung „Kindermuseum“ treffend.



Abbildungen 10 bis 17: Eindrücke aus dem Besuch des Kindermuseums.

Das „Naturkundemuseum“ besteht aus einem über 20 Meter langen Gang, der zu beiden Seiten Dioramen, Schaukästen und Schautafeln zu ökologischen Inhalten bietet (vgl. Abbildung 18). Die zur Charakterisierung verwendeten Begriffe werden wie folgt verwendet: Diorama: plastische Darstellung mit gemaltem Hintergrund. Schaukasten: Ausstellungsvitrine. Schautafel: zweidimensionale bildliche Darstellung zu einem bestimmten Thema.



Abbildung 18: Naturkundemuseum.

Im Lindenhof sind Dioramen stilisierte „Standbilder“ realer Natur, die sehr viele makroskopisch sichtbare Tiere pro Fläche zeigen (z. T. weit über 30 Tiere auf unter zehn Quadratmetern) und in manchen Fällen durch technische Vorrichtungen inszenierte Begegnungen mit Naturphänomenen, z.B. mit Tierstimmen, bieten können. Der Vordergrund wird durch dreidimensionale Objekte, Originale oder nachgeahmte Repräsentationen, gestaltet. Den Hintergrund bildet ein möglichst authentisches Landschaftsbild, das in den Vordergrund übergeht, so dass ein „natürlicher“ Gesamteindruck entsteht. Schaukästen sind räumlich abgegrenzte dreidimensionale Darstellungen zu einem bestimmten Thema, wobei die Ausgestaltung sehr variabel ist. Manche Schaukästen wirken eher spartanisch, andere sind durch viele dreidimensionale Strukturelemente gestaltet. Die Schautafeln liefern in Abbildungen, Tabellen und Texten ergänzende Informationen zu den Schaufenstern (Dioramen, Schaukästen).

Die rein formale Gliederung des „Naturkundemuseums“ bedient sich hauptsächlich zweier Gliederungsebenen: Themenkreise (im Museum rot unterlegt) und „Unterthemen“ (im Museum blau unterlegt). Unterstreichungen deuten die spätere Verwendung als Arbeitsstation für die vorliegende empirische Untersuchung an.

Themenkreise und Unterthemen des „Naturkundemuseums“:

Der Wald – Naturlandschaft in Mitteleuropa

Was wäre die Landschaft ohne Mensch (Schautafel)

Urwald im Fichtelgebirge (Diorama)

Das Mosaik-Zyklus-Konzept (Schautafel)

Der Mensch nutzt

Jungsteinzeit (Diorama)

18. Jahrhundert (Diorama)

Gegenwart (Diorama)

Artenreichtum aus Mangel

Ursachen des Mangels (Schautafel)

Wacholderheide im Frankenjura (Diorama)

Von der Artenvielfalt (Schaukasten)

Feuchtgebiete

Ökosystem Weiher (Schautafel)

Ökosystem Weiher (Diorama)

Artenvielfalt der Kulturlandschaft

Vielfalt durch Eingriffe (Schautafel)

Steinkauz und Brachvogel (Schaukasten)

Vom Kulturfolger zum Kulturflüchter

Der Wald in der Landwirtschaft (Schautafel)

Das Schicksal von Kornblume, Feldhase und Rebhuhn (Schaukasten)

Lebensraum aus zweiter Hand

Steinbruch (Schaukasten)

Sandgrube in Nordbayern (Diorama)

Kiesgrube (Schautafel)

Lebende Messinstrumente

Vögel und Pestizide (Schaukasten)

Was ist ein Bioindikator (Schautafel)

Was will Naturschutz?

Wie viel Platz für Tiere? (Schaukasten)

Was braucht ein Falke zum Leben? (Schaukasten)

Ziele (Schautafel)

Maßnahmen (Schautafel)

Bereits ausgestorben... (Schaukasten)

Grundthema dieses Museumsteils ist die Beeinflussung der „Natur“ durch den Menschen. Die Umsetzung sei im Folgenden kurz skizziert: Beginnend mit dem Diorama „Urwald im Fichtelgebirge“ (Abbildung 19) wird ein Urzustand von unbeeinflusster Natur dargestellt. Der Einfluss des Menschen nimmt zu. Durch menschliche Besiedlung verändert sich die Fränkische Schweiz, bis Biotope entstehen, die sich ohne menschliche Nutzung sehr schnell linear verändern würden wie die „Wacholderheide im Frankenjura“ (Abbildung 20; Sekundärsukzession, vgl. Kalusche 1999, S. 96 ff.). Dieses Diorama liegt etwa auf halber Höhe des rechten Ganges. Am Ende des Ganges wird die Darstellung des oft schädlichen Einflusses des Menschen (auf der linken Seite, ganz hinten) mit dem Schaukasten „Das Schicksal von Kornblume, Feldhase und Rebhuhn“, in der Untersuchung als „Kulturfolger – Kulturflüchter“ bezeichnet (Abbildung 23), fortgeführt. Neue Biotope, „Lebensräume aus zweiter Hand“, entstehen, z.B. ein „Steinbruch“ – hier wird das Thema „Archaeopteryx“ behandelt (Abbildung 24) - oder eine „Sandgrube“ (Abbildung 21, 22), die etwa auf halber Höhe des linken Ganges liegt.



Abbildungen 19 bis 25: Die „Schaufenster“: Die Bezeichnungen richten sich nach den in der Untersuchung verwendeten Begriffen, nicht notwendigerweise nach den Benennungen im Museum (vgl. Erläuterungen unten; Museumsbezeichnung in Klammern). Benennung von oben links beginnend: „Urwald“ (Urwald im Fichtelgebirge), „Wacholderheide“ (Wacholderheide im Frankenjura), „Sandgrube“ (Sandgrube in Nordbayern), Teil der Legende zur „Wacholderheide“, „Kulturfolger – Kulturflüchter“ (Das Schicksal von Kornblume, Feldhase und Rebhuhn), „Archaeopteryx“ (Steinbruch), „Platzbedarf“ (Wie viel Platz für Tiere?).



Weitere Probleme werden erörtert, z.B. der Verbrauch von Naturräumen durch Menschen. In diesem Zusammenhang wird der Raumbedarf unterschiedlicher Vögel u. a. anhand des Schaukastens „Platzbedarf“ thematisiert (Abbildung 25). Der vorderste Schaukasten links, „Bereits ausgestorben ...“, stellt, als eine mögliche besonders negative Folge menschlicher Eingriffe, in Bayern ausgerottete Vögel vor.

Die unterstrichenen Dioramen bzw. Schaukästen entsprechen dem oben dargestellten Grundgedanken des Museums in repräsentativer Art und Weise. Sie wurden für den geplanten Museumsbesuch gemäß folgender Kriterien ausgewählt: Für Schüler im Alter von zehn Jahren sollte eine sinnvolle Beschäftigung von etwa zehn Minuten Dauer mit jedem der „Schaufenster“ möglich sein, auch ohne zu sehr Details behandeln zu müssen. Die Hauptaussage sollte demnach nicht zu einfach sein, wie z.B. im Schaukasten „Bereits ausgestorben...“. Diese Tatsache an sich gibt dem Schüler in der Situation des Museumsbesuchs zu wenige Handlungsoptionen. Gleichzeitig darf die Hauptaussage nicht zu komplex sein. In einem Schaukasten wird beispielsweise die Anreicherung von DDT (Gesarol bzw. 1,1-Di(4-chlorphenyl)-2,2,2-trichlorethan, vgl. Beyer & Walter, 1988, S. 605) in der Nahrungskette thematisiert. Bei dem Endkonsumenten - als Beispiele werden aufgeführt Wanderfalke und Waldkauz - führt diese Akkumulation zu verringerter Nachkommenzahl. Das Thema (und die Darstellung) dieses Schaukastens sind für Schüler der fünften Jahrgangsstufe wahrscheinlich zu anspruchsvoll. Man kann z.B. nicht davon ausgehen, dass alle Schüler das schon didaktisch stark reduzierte Konzept „Nahrungskette“ verinnerlicht haben. Ohne dieses Minimalwissen ist jedoch der Hauptgedanke nicht verstehbar. Eine andere Quelle von Unverständlichkeit kann eine zu schlechte didaktische Aufbereitung eines Themas sein. In dem Schaukasten „Was braucht ein Falke zum Leben?“ wird der Revierbedarf eines Vogels von globaler Perspektive ausgehend und sich dem Einzelrevier sukzessive nähernd beschrieben. Ohne näher auf Details einzugehen, ist hier die Darstellung der Hauptaussage so missverständlich, dass sie auch für Studenten, die diesen Schaukasten im Rahmen von Lehrveranstaltungen bearbeiteten, nicht zu begreifen war. Eine Veranschaulichung, die offensichtlich so wenig selbsterklärend ist, kann mit Rücksicht auf ein noch zu präzisierendes weitgehend selbstgesteuertes Treatment nicht in die Auswahl der sechs zu bearbeitenden „Schaufenster“ aufgenommen werden. Zu große Detailfülle und Artenvielfalt, wie z.B. in dem Diorama „Weiher“, ist ebenfalls nicht angemessen. Als weiteres eher technisches Kriterium

werden die „Schaufenster“ nach der Maßgabe ausgewählt, dass zu ähnliche Stationen, z.B. zwei idealisierende Biotope darstellende Dioramen, sich nicht direkt nebeneinander befinden sollten, um ungewollte Zusammenarbeit zweier Schülergruppen zu vermeiden. Ferner sollte der Charakter des Museums in der Auswahl möglichst erhalten bleiben.

Die drei biotopabbildenden und die drei themenbezogenen Stationen (siehe Unterstreichungen oben) wurden gemäß der obigen Vorgaben ausgewählt.

Aus fachdidaktischer Sicht ist ein Naturkundemuseum ein außerschulischer Lernort wie auch ein Zoo oder ein Botanischer Garten, der inhaltliche und methodische Alternativen zum Unterricht im Klassenraum bieten kann (vgl. Killermann 1999, S. 11). Das bezieht sich nicht nur auf die kognitive Ebene berührende unterrichtliche Möglichkeiten sondern auch solche, die die affektive Ebene ansprechen. „(...) die Naturkundeabteilung innerhalb eines [entsprechend großen] Museums [besitzt] für Schüler bei Weitem die stärkste Anziehungskraft“ (Killermann 1995; S. 238). Gries (1996, S. 8) bezeichnet Naturkundemuseen als „Erlebnis- und Lernorte“. In „Naturkundemuseum“ und „Kindermuseum“ des Lindenhofs werden den Besuchern originale Begegnungen und z. T. multisensorisches Erleben geboten, so dass affektive Erfahrungen möglich werden, so wie es außerschulischen Lernorten im besonderen Maße entspricht (Killermann, 1995, 1999).

1.2.3. Fachliche Charakterisierung

Folgende Vorgehensweise bietet sich an: Die fachlichen Inhalte des Museums werden kurz beschrieben und analysiert. Ziel ist es, zu beurteilen, ob die Darstellungen fachlich (und fachdidaktisch) im Hinblick auf den Museumsbesuch akzeptabel sind.



Abbildungen 26, 27: Schüler bei der Bearbeitung von „Urwald im Fichtelgebirge“.

a) Das Diorama „Urwald im Fichtelgebirge“ (vgl. Abbildungen 26, 27):

Das Fichtelgebirge befindet sich im Nordosten Bayerns. Die Bezeichnung „Fichtelgebirge“ leitet sich aus dem hauptsächlichen Vorkommen von Fichten in dieser Region her (Küster 1995, S. 65). Bezieht man sich auf Ellenbergs (1996, S. 23) Darstellung „der Vegetation Mitteleuropas (...) um Christi Geburt, d.h. vor stärkeren Eingriffen des Menschen“, wie man die Bezeichnung „Urwald im Fichtelgebirge“ interpretieren kann, so liegt das Fichtelgebirge in der Vegetationszone folgender Beschreibung: „Buchenwald-Berglagen mit Tanne und (oder) Fichte (...) = subalpiner Buchenwald“. Dieser Darstellung entspricht das Diorama. Links im Vordergrund ist eine Rotbuche (*Fagus sylvatica*) dargestellt. In der bildlichen Weiterführung des plastisch ausgestalteten Vordergrunds ist ein Buchenmischwald auf hügeligem Untergrund zu erkennen. Die Buchen werden von Inseln von Fichten durchbrochen. Der Vordergrund ist gestaltet durch konservierte natürliche oder naturgetreu nachgeahmte Pflanzen, Totholz, andere verrottende Pflanzenteile, Pilze, Moose, Flechten und v. a. durch Tiere. Eine genauere Beschreibung der mit Hilfe von plastischen Modellen oder zweidimensional abgebildeten Pflanzen ist z. T. aufgrund ihrer schlechten Qualität problematisch und wenig hilfreich. Diese v.a. pflanzlichen Abbilder und Modelle sollten für die Schüler möglichst wenig genutzt werden. Sie sind eher als „einigermaßen“ authentische Kulisse zu sehen, nicht aber als Anregung zu fachlicher Auseinandersetzung. Selbst die oben erwähnte Rotbuche ist nur mit Mühe zu identifizieren, insbesondere wegen der schlechten Nachahmung der Blätter. Bei Tieren sind die Probleme grundsätzlich geringer, da bei Vögeln und Säugetieren ausgestopfte Originale Verwendung finden. Bei Insekten stellt man ebenfalls das präparierte originale Tier aus. Von Reptilien und Amphibien sind haltbare, sehr authentische künstliche Präparate vorhanden (vgl. Somso Modelle /Firmenkatalog 2003, S. 30 f.; beachte z.B. Laubfrösche S. 31.).

Die Legende führt folgende mit Hilfe des Schaubildes identifizierbare Arten auf:

1. Nagelfleck (*Agria tau*)
2. Kaisermantel (*Argynnis paphia*)
3. C-Falter (*Polygonia c-album*)
4. Lederlaufkäfer (*Carabus coriaceus*)
5. Dungkäfer (*Scarabaeidae*)
6. Feuersalamander (*Salamandra salamandra*)
7. Waldschnepfe (*Scolopax rusticola*)
8. Tannenhäher (*Nucifraga caryocatactes*)
9. Rotkehlchen (*Erithacus rubecula*)
10. Buchfink (*Fringilla coelebs*)
11. Buntspecht (*Dendrocopos major*)
12. Schwarzspecht (*Dryocopus martius*)
13. Tannenmeise (*Parus ater*)
14. Sperlingskauz (*Glaucidium passerinum*)
15. Haselhuhn (*Tetrastes bonasia*)
16. Grünspecht (*Picus viridis*, ergänzt)

Bezüglich zur Fauna wurde anhand von Grzimeks Tierleben, Brohmer (1984), Bauer (1997), Jacobs & Renner (1988), Kudrna, (2002), Bezzel (1982), Nitsche & Plachter (1987), Gubitz & Pfeifer (1993), Grant & Zetterström (2000) und Ergänzungen von Kalusche (1999) überprüft, inwiefern die dargestellten Tiere dem beschriebenen Biotop entsprechen können.

An Insekten ist zuerst die einheimische Schmetterlingsart „Nagelfleck“ (*Agria tau*) genannt, die man an Rotbuchen findet (Dierl 1969, S. 340). Kaisermantel (*Argynnis paphia*) und C-Falter (*Polygonia c-album*) gehören zur Familie der Fleckenfalter (*Nymphalidae*), beides heimische Arten (Dierl 1969, S. 366 f.). Kaisermantel findet man auf Waldwiesen (Jacobs & Renner 1988, S. 443). Lederlaufkäfer (*Carabus coriaceus*) sind heimische Vertreter der Familie Carabidae (Laufkäfer) (Jacobs & Renner 1988). Das Präparat des in der Legende ausgewiesenen Dungkäfers ist nicht (mehr) vorhanden. Als einziger Lurch wird der Feuersalamander dargestellt. Es handelt sich um die östliche Verbreitungsform, Gefleckter Feuersalamander (*Salamandra salamandra salamandra*) (Freytag 1970, S. 329, Diesener & Reichholf 1996, S. 12 ff.). Es

folgt die Klasse der Vögel (Aves). Bei ihnen sind z.T. Möglichkeiten vorgesehen, Rufe oder Gesang anzuhören. Da ein Abrufen von Lauten bei der Bearbeitung der Stationen aus organisatorischen Gründen unterbleiben muss, wird auf Gesang und Rufe nicht weiter eingegangen. Die Waldschnepfe (*Scolopax rusticola*) ist ein typischer Waldbewohner, der der Avifauna von Fichtenwäldern und anderen Waldtypen angehören kann (Bezzel 1982, S. 243, 251). Sie ist in Wäldern des Fichtelgebirges nachgewiesen (Gubitz & Pfeifer 1993, S. 187 ff.). Tannenhäher, präzise Eurasiatische Tannenhäher (*Nucifraga caryocatactes*), bevorzugen Nadelwälder vorwiegend gebirgiger Regionen; ihr Hauptverbreitungsgebiet liegt in Asien (Wüst 1979, S. 494). Im Fichtelgebirge sind sie mäßig häufig (Gubitz & Pfeifer 1993, S. 456 ff.). Das Rotkehlchen (*Erithacus rubecula*) ist fast überall in Europa und in Teilen Asiens verbreitet (Dorst 1970, S. 271 f.) und in Buchen- und Fichtenbeständen häufig, wie auch der Buchfink (*Fringilla coelebs*) (Bezzel 1982, S. 247 und Newton 1970, S. 387 f.). Die Familie der Spechte (*Picidae*) sind vertreten durch den in Europa und Asien beheimateten Buntspecht (*Dendrocopos major*), den in Mittel- und Nordeuropa bis Japan verbreiteten Schwarzspecht (*Dryocopus martius*) und den v.a. in Europa vorkommenden Grünspecht (*Picus viridis*) (Wendt 1970, S. 105 ff.). Der Buntspecht bevorzugt Laub- und Mischwald vor reinem Nadelwald, wo er durchaus auch zu finden ist (Bauer 1997, S. 289); Schwarzspechte dagegen sind im Nadelwald häufiger als im Mischwald (Bauer 1997, S. 287), während Grünspechte „vorzugsweise am Rand geschlossener Bestände“ von Mischwald, seltener von Nadelwald auftreten (Bezzel 1982, S. 244). Buntspecht und Schwarzspecht sind in Ostoberfranken „mäßig häufig“, der Grünspecht ist „ein nicht häufiger Jahresvogel“ (Gubitz & Pfeifer 1993, S. 284 ff.; vgl. Nitsche & Plachter 1987, S. 138 ff.). Spechte stellen in mindestens zweifacher Hinsicht günstige Beispiele bezüglich des Begriffs „Ökologische Nische“ dar: Zum einen durch ihre offensichtlichen Angepasstheiten mit Meißelschnabel und Harpunenzunge, Kletterfuß und Stützschwanz an eine bestimmte Art der Nahrungssuche bzw. Lebensweise, wodurch diese Gruppe „eine“ leicht einsehbare ökologische Nische einnimmt; zum anderen gibt es jedoch im selben Verbreitungsgebiet mehr als eine Art dieser Tiergruppe, z.B. leben in Mitteleuropa mindestens acht Spechtarten. Durch diese Koexistenz wird die scheinbar eine Nische in acht unterschiedliche Nischen aufgespalten. Diese differenziertere Einnischung wird möglich aufgrund „kleinster Abweichungen im Verhalten“ der unterschiedlichen Arten (Kalusche 1999, S. 84 ff.). Die Tannenmeise (*Parus ater*), seltener in Buchen-, aber häufig in Kiefern- und

sehr häufig in Fichtenwäldern vorkommend, ist die kleinste heimische Meisenart (Berndt 1970, S. 300 f.; Bauer 1997, S. 415 ff.). Sperlingskäuze (*Glaucidium passerinum*) sind typische Bewohner von Nadelwäldern, deren Verbreitungsgebiet v.a. Nordeuropa und Asien ist und sich gerade noch bis Mitteleuropa erstreckt (König 1969, S. 397 ff.; Bauer 1997, S. 252 ff.). Sie gelten zwar insgesamt als selten, kommen jedoch „im Fichtelgebirge öfters vor“ (Gubitz & Pfeifer 1993, S. 239). Dieses Diorama „Urwald im Fichtelgebirge“ stellt mit Totholz und angedeuteten kleinen Lichtungen in nahezu idealer Weise ein Haselhuhn-Biotop dar (vgl. Bezzel 1982, S. 243; Bauer 1997, S. 129 f.). Haselhühner (*Tetrastes bonasia*), reine Waldbewohner, sind verbreitet in Asien, Nord- und Mitteleuropa (Müller-Using 1968, S. 459). In Ostoberfranken sind sie sehr selten geworden (Gubitz & Pfeifer 1993, S. 134 ff.).

Viele Stopfpräparate von Vögeln und Säugetieren sind leicht beschädigt. Besonders deutliche Schäden haben die präparierten Insekten. Einige sind anhand von Legende und Schaubild gar nicht zu finden, oder sind so weit zerstört, dass eine eindeutige Identifizierung für die Museumsbesucher nicht möglich ist.

Insgesamt lässt sich auf der Ebene dieser Recherche für das Diorama „Urwald im Fichtelgebirge“ kein fachlicher Fehler nachweisen. Die Erstellung des Dioramas erfolgte solide. Es kann nicht beurteilt werden, ob in der Konzeption des Dioramas wesentliche Arten fehlen oder bezüglich der wahrscheinlich dargestellten Zeit und der recherchierten Fauna kleinere Inkonsistenzen bestehen, denn diese grobe Überprüfung der Vegetation bezieht sich auf die Zeit Christi, die Angaben zur Verbreitung der Tiere auf die Gegenwart. Da jedoch keine größeren klimatischen Verschiebungen in diesem „kurzen“ Zeitraum von zwei Jahrtausenden stattfanden, werden die für dieses Diorama in Kauf genommenen Unschärfen für vertretbar gehalten. Der aktuelle Zustand des Dioramas insbesondere der einiger Stopfpräparate und der der Insektenpräparate ist nicht optimal. Insbesondere jegliche auf Insekten bezogene Frage wird daher in den Arbeitspaketen für die Schüler vermieden.



Abbildungen 28, 29: Schüler bei der Bearbeitung von „Wacholderheide“.

b) Das Diorama „Wacholderheide im Frankenjura“ (vgl. Abbildung 28, 29):

Vor der Beschreibung des Dioramas erfolgt eine kurze Begriffsklärung und Schilderung der Entstehung der „Wacholderheiden“. Das trockene, warme Klima lichter Eichenwälder begünstigt die im Vergleich zum Fichtelgebirge frühere Besiedlung der Fränkischen Schweiz etwa ab 5000 vor Christus (Weid 1995, S. 4). Die menschliche Nutzung, Rodungen und v. a. Schafhut, drängt den Wald so weit zurück, dass in der nun entstandenen Kulturlandschaft Mitte des 19. Jahrhunderts nur noch ca. 20% der Fläche der Fränkischen Schweiz mit Wald bedeckt sind (Weid 1995, S. 4). Nach Weid (1996, S. 3 f.) ist „Wacholderheide“ die im Volksmund verwendete Bezeichnung der beweideten oder ehemals beweideten Magerrasen bzw. (Halb-)Trockenrasen mit Wacholderbüschen und anderen Sträuchern und Bäumen auf meist kalkhaltigen Böden des nördlichen Ausläufers der Fränkischen Alb (vgl. Küster 1995, S. 22 f.): der Fränkischen Schweiz. „Die Magerkeit ist durch den einseitigen Nährstoffentzug der Beweidung zustande gekommen, weil die Schafe v.a. in der Nacht im Pferch, der außerhalb der Magerrasen auf einem Acker liegt, abkoten“ (Weid 1996, S. 3; vgl. Ellenberg 1996, S. 50 f.). Die Pflanzengesellschaft dieser Magerrasen hat sich an die besondere Nährstoffarmut, die Wasserarmut und extreme Temperaturen in den Sommermonaten angepasst, z.B. indem sie „das Dreifache der oberirdischen Biomasse in den Wurzeln [anlegen]“ (Weid 1996, S. 9). In der Gegenwart verändert sich das Bild der Fränkischen Schweiz erneut. Durch verminderte Beweidung der letzten Jahrzehnte beginnen viele der Heiden auf kalkhaltigem Untergrund zu verbuschen (vgl. Weid 1995, S. 18 ff.). Durch kontrollierte Rodungsaktionen / Entbuschungen versucht man dem zu begegnen.

Zum Diorama: In der den Hintergrund bildenden Landschaftsdarstellung wechseln sich freie Heidefläche mit vereinzelt Wacholdersträuchern oder anderen holzigen Gewächsen mit recht stark verbuschten Bereichen ab. Zu sehen sind Erhebungen, Abhänge und stark verwitterte Kalkfelsformationen, die sich im Vordergrund wiederfinden. Kleine Sträucher (z.B. Wacholder, Heckenrose), verdorrte Grasbüschel und eine Vielzahl von Tieren, darunter besonders viele Insekten, sind v. a. in Form von präparierten Originalen ausgestellt.

Die Legende führt folgende mit Hilfe des Schaubildes identifizierbare Arten auf:

1. Segelfalter (*Iphiclides podalirius*)
2. Apollofalter (*Parnassius appollo*)
3. Hauhechelbläuling (*Polyommatus icarus*)
4. Schwalbenschwanz (*Papilio machaon*)
5. Blutströpfchen (*Zygaena filipendulae*)
6. Rotflügelige Schnarrschrecke (*Psophus stridulus*)
7. Gemeine Sichelschrecke (*Phaneroptera falcata*)
8. Goldlaufkäfer (*Carabus auratus*)
9. Wespenbock (Großer Wespenbock = *Necydalis major*)
10. Zauneidechse (*Lacerta agilis*)
11. Heideschnecke (Familie Helicaceae, wahrscheinlich *Xerolenta spec.*)
12. Wanderfalke (*Falco peregrinus*)
13. Feldgrille (*Gryllus campestris*)
14. Neuntöter (*Lanius collurio*)
15. Raubwürger (*Lanius excubitor*)
16. Hausrotschwanz (*Phoenicurus ochruros*)
17. Steinschmätzer (*Oenanthe oenanthe*)

Weid (1996, S. 17 ff.) hebt als besonders typische Arten der Fränkischen Schweiz z.B. Segelfalter und Apollofalter, Gemeine Sichelschrecke, Rotflügelige Schnarrschrecke und Neuntöter hervor. Alle vorkommenden Insekten sind in Jakobs & Renner (1988) als mitteleuropäische Arten identifiziert. Der Hauhechelbläuling ist überall in Deutschland häufig (Kudrna 2002, S.265) Die Fraßpflanzen z.B. Hauhechel und Klee sind in der Fränkischen Schweiz häufig. Das Blutströpfchen gilt als häufigste Art der Zygaenidae und ist mit Ausnahme des hohen Nordens und Südspaniens in fast

ganz Europa beheimatet (<http://www.altmuehltal.de/greding/lehrpfad/blutstroepfchen.htm>). Der Schwalbenschwanz ist in ganz Deutschland verbreitet (Kudrna 2002, S.230). Zauneidechsen sind in Deutschland mit der Ausnahme höherer Gebirgslagen häufig (Brohmer 1984, S. 453). Das Vorkommen der Heideschnecke an trockenen Grashängen wird beschrieben, insbesondere das im Diorama dargestellte Massenvorkommen in Trockenruhe (Fechter & Falkner 1989, S. 208). Allein Vögel seien etwas genauer betrachtet. Laut Gubitz & Pfeifer (1993, S. 131 ff.) sind Wanderfalken seltene Brutvögel in der Fränkischen Schweiz. Es gibt zahlreiche Nachweise aus der ersten Hälfte des 20. Jahrhunderts, ab 1950 nimmt die Anzahl besetzter Horste stetig ab, bis ab 1970 praktisch keine Nachweise für brütende Wanderfalken mehr existieren. 1992 gelang im Landkreis Forchheim nach einer oberfrankenweiten Auswilderungsaktion wieder die erfolgreiche Brut eines Paares. Die beiden Würger (Familie Laniidae) Neuntöter (selten bis mäßig häufig) und Raubwürger (sehr selten) sind als Brutvögel des zu untersuchenden Gebiets nachzuweisen (Gubitz & Pfeifer 1993, S. 439 ff. und S. 444). Hausrotschwänze sind häufige Brutvögel (Gubitz & Pfeifer 1993, S. 358 ff.; Bauer 1997, S. 336 f.). Von Steinschmätzern gibt es nur sehr wenige Brutnachweise in der nördlichen Frankenalb (Gubitz & Pfeifer 1993, S. 373 ff.).



Abbildungen 30, 31: Schüler bei der Bearbeitung von „Sandgrube in Nordbayern.“



c) Das Diorama „Sandgrube in Nordbayern“ (vgl. Abbildung 30, 31):

Durch den Abbau von Rohstoffen, z.B. Sand, verändern sich Vegetation und Landschaft (Küster 1995, S. 263 ff.) Sekundärbiotope entstehen. Bezzel (1982, S. 22 ff.) benennt einige dieser „Biotope der Kulturlandschaft“, z.B. „Ruderalfluren“, „Brach-

land, Kiesgruben“, Abgrabungsgewässer (Sand-, Kiegruben). Die Legende führt für das Diorama „Sandgrube in Nordbayern“ folgende mit Hilfe des Schaubildes identifizierbare Arten auf:

1. Feld-/ Sandlaufkäfer (*Cicindela campestris*)
2. Großer Blaupfeil (*Orthetrum cancellatum*)
3. Plattbauch (*Libellula depressa*)
4. Sandwespe (*Ammophila spec.*)
5. Maulwurfgrille (*Gryllotalpa gryllotalpa*)
6. Admiral (*Vanessa atalanta*)
7. Flussregenpfeifer (*Charadrius dubius*)
8. Hermelin (*Mustela erminea*)
9. Wiedehopf (*Upupa epops*)
10. Feldsperrling (*Passer montanus*)
11. Eisvogel (*Alcedo atthis*)
12. Bienenfresser (*Merops apiaster*)
13. Kreuzkröte (*Bufo calamita*)
14. Gelbbauchunke (*Bombina variegata*)

Alle genannten Insekten sind mitteleuropäische Arten (Jakobs & Renner 1988). Laut Legende sollten sie im Diorama zu finden sein. Dies ist nicht der Fall. Viele sind stark beschädigt oder fehlend. Sie können von den Schülern nicht genutzt werden. Eine präzisere Analyse ist darum überflüssig. Zwei Amphibien sind genannt und im Diorama zu sehen. Die Kreuzkröte bevorzugt sandigen Untergrund (Brohmer 1984, S. 449). Ein Vorkommen in Sandgruben ist wahrscheinlich. Gelbbauchunken finden sich in Gräben, Pfützen und kleinen Teichen, Biotopen wie in der dargestellten Sandgrube (Brohmer 1984, S. 448). Hermeline, die einzigen Säuger des Dioramas, sind in ihrer Biotopwahl sehr anpassungsfähig und sind „allgemein gern (...) in wassernahen Biotopen“ (Corbet & Ovenden 1980, S. 186). Ein Vorkommen in Sandgruben, mit kleinen stehenden Gewässern, etwas Unterholz und insgesamt eher offenem Gelände ist demnach nicht auszuschließen. Zu den Vögeln: Flussregenpfeifer sind seltene Brutvögel (Gubitz & Pfeifer 1993, S. 169 ff.). Wenn sie in Nordbayern nisten, dann in Biotopen wie Sandgruben (Gubitz & Pfeifer 1993, S. 172). Als „Offenlandbrüter“ (Bezzel 1982, S. 49) kann ein Wiedehopf sicher in verlassenen Sandgruben gefunden werden. Seine Seltenheit in Nordbayern - seit ca. 1960 brüten Wiedehopfe nicht

mehr in Ostoberfranken - macht es unwahrscheinlich, ihn zu sehen; es gibt nur wenige Nachweise von Sichtungen (Gubitz & Pfeifer 1993, 271 ff.). Ebenfalls äußerst selten ist der Bienenfresser. Es gibt einige wahrscheinliche Sichtungen in Franken (Gubitz & Pfeifer 1993, S. 273), Nitsche & Plachter (1987, S. 249) nennen wenige Brutpaare in Bayern. Aufgrund von Klimaveränderungen sollte der Bestand in Mitteleuropa eher zunehmen (Bezzel 1982, S. 67). Offene Gelände mit Steilhängen, wie man sie in Sandgruben nicht selten finden kann, sind typische Biotope für Bienenfresser (Grant, Zetterström 2000, S. 222). Die Wahrscheinlichkeit einer Sichtung ist sehr gering. Eisevögel dagegen sind etwas häufiger. Sie brüten regelmäßig in Nordbayern (Gubitz & Pfeifer 1993, S. 266 ff.). Ihre Nisthöhlen finden sich in „Böschungen und Steilufern von Gewässern oder Abbaustellen“ (Gubitz & Pfeifer 1993, S. 271). Feldsperlinge gelten als mäßig häufig (Gubitz & Pfeifer 1993, S. 478 f.).

Ob man durch das (vielleicht sehr seltene) Vorkommen attraktiver Vögel in stillgelegten Sandgruben (Wiedehopf, Eisevogel und Bienenfresser) für Schüler den Wert dieser Lebensräume verdeutlichen sollte oder ob dadurch falsche Vorstellungen geweckt werden, soll hier nicht entschieden werden. Die museale Vorgabe wird aus fachlicher (und fachdidaktischer Sicht) für vertretbar gehalten.



Abbildungen 32, 33: Schüler bei der Bearbeitung von „Kulturfolger – Kulturflüchter“.

d) Der Schaukasten „Kulturfolger - Kulturflüchter“ (vgl. Abbildung 32, 33):

Im diesem Schaukasten geht es um „Das Schicksal von Kornblume, Feldhase und Rebhuhn.“ Als das zentrale Thema wird begriffen, wie aus Kulturfolgern Kulturflüchter werden. In dieser Untersuchung ist es darum als „Kulturfolger - Kulturflüchter“ bezeichnet. Die Ursache für die oben angedeutete Entwicklung liegt in der gravieren-

den Veränderung der Kulturlandschaft in den letzten Jahrzehnten. „In unseren Kulturlandschaften (geographisch sind Mitteleuropa und Randlandschaften gemeint; vgl. Bezzel 1982, S. 10) ist der Mensch der wichtigste faunenbildende Faktor geworden (...). Vorgänge, die (...) in geologischen Zeiträumen abliefen, (...) [werden] auf Jahrzehnte, ja sogar Jahre zusammengedrängt (...). Als Folge (...) [befindet] die Fauna sich in einem bisher nicht dagewesenen Umwälzungsprozess“ (Bezzel 1982, S. 8 f.). Während die Landschaft vom Mittelalter bis zum 20. Jahrhundert eher die Artenvielfalt gefördert hat, „verkehrt sich diese Entwicklung [im 20. Jahrhundert] ins Gegenteil (Bezzel 1982, S. 19). Betrachtet man die „freilebende Tierwelt in Mitteleuropa“, so sind – „je nach untersuchter [Tier-]gruppe“ – etwa ein bis zwei Drittel des Artenbestandes gefährdet (Reichholf 1988, S. 272). Ursache ist v.a. die Industrialisierung der Landwirtschaft. Den Beginn dieser gravierenden Veränderungen in Mitteleuropa markiert das Nationalsozialistische Regime. „(...) die ‚Erzeugungsschlacht‘ sollte Deutschland auf dem Gebiet der Nahrungsversorgung autark machen“ (Küster 1995, S. 354). Vereinzelt gibt es frühere Nachweise des Rückgangs der Populationsdichte von Kulturfolgern wie z.B. beim Rebhuhn (*Perdix perdix*). Aus Ostoberfranken sind von 1932 Berichte bekannt über „negative [Bestands-]veränderungen im Brutgebiet durch die Intensivierung der Landwirtschaft“ (Gubitz & Pfeifer 1993, S. 153). Stellvertretend für die im Naturkundemuseum des Lindenhofs dargestellten Kulturfolger (z.B. Goldammer, Rebhuhn, Stieglitz, etc.), die in der zweiten Hälfte des 20. Jahrhundert Kulturflüchter werden, sei der Feldhase (*Lepus europaeus*) herausgegriffen. Feldhasen sind in ganz Deutschland verbreitet (Brohmer 1984, S. 522). Sie bevorzugen offene Landschaften mit vielfältiger Bodennutzung und großer Vegetationsvielfalt (Seck-Lanzendorf 1997, S. 5 f.). Die langsame Veränderung der Kulturlandschaft fast bis zur Mitte des 20. Jahrhunderts erlaubte es dem Feldhasen, sich anzupassen; erst dann beschleunigt sich der Strukturwandel der Landwirtschaft. Heute ist „die Populationsstruktur seines Besatzes weitgehend zerrüttet“ (Seck-Lanzendorf 1997, S. 11). Entscheidend für die Dichte der Population ist die Strukturierung der Landschaft, v. a. die Größe der einzelnen Felder, die Anzahl von Hecken, Gebüsch und Wegrändern (Seck-Lanzendorf 1997, S. 12). Weiter – so Seck-Lanzendorf (1997, S. 14 ff.) – bestimmen Schadstoffe, toxisch wirksame Substanzen, in Luft, Wasser, Boden und Futter die Populationsentwicklung. Zu Schadstoffen zählen z.B. Pestizide (Herbizide, Insektizide, Fungizide), die indirekt die Population beeinflussen, weil Futterpflanzen der Hasen geschädigt werden, die aber auch direkt durch Intoxikation wirken können.

Weitere Schadstoffe sind z.B. Düngemittel und Schwermetalle (Seck-Lanzendorf 1997, S. 14 ff.). Direkte Wirkung auf den Bestand haben die Zunahme des Straßenverkehrs, Jagd, Prädatoren wie Fuchs (*Vulpes vulpes*), Marder (Fam. Mustelidae, bis hin zum Mauswiesel (*Mustela nivalis*)) und Greifvögel (z.B. Mäusebussard (*Buteo buteo*)) außerdem Infektionskrankheiten (Seck-Lanzendorf 1997, S. 14 und 18 ff.). Einige dieser direkten Einflüsse sind im 20. Jahrhundert sicher nicht verschärft worden, z.B. der Druck durch Prädatoren. Beutegreifer kontrollieren i.d.R. nicht die Population von Beutetieren (Bay & Rodi 1991, S. 32 ff.). Ihr Einfluss auf eine deutlich geschwächte Population kann jedoch andere Folgen haben, als auf eine „gesunde“ (Bay & Rodi 1991, S. 35).

Bei pflanzlichen Beispielen in dem oben genannten Schaukasten, z.B. der Kornblume, ist wieder die schon angesprochene Schwierigkeit der fehlenden Detailgetreue dieser Präparate anzumerken. Pflanzen werden darum weder bei der fachlichen Charakterisierung noch bezüglich des Unterrichts besonders berücksichtigt. Ferner sind Fotos einer typischen Monokultur, des Weiteren einer ungemähten Wiese dominiert von blühendem Mohn (*Papaver spec.*) – nach Schmidt (1979, S. 90) könnte man von einem „Mohn-Aspekt“ sprechen - und einer Böschung mit Feldkräutern und kleinen Büschen ausgestellt. Präparate weitere Gräser, Feldkräuter und einer Getreidegarbe vermitteln neben den Stopfpräparaten von Feldhase (*Lepus europaeus*), Rebhuhn (*Perdix perdix*), Goldammer (*Emberiza citrinella*) und Stieglitz (*Carduelis carduelis*) den Eindruck von Authentizität.

Das komplexe Thema wird in dem Schaukasten fachlich und didaktisch akzeptabel behandelt, wenngleich erst textliche Hilfen und nicht die Exponate für sich selbst stehend dem Museumsbesucher entscheidende Zusammenhänge des Themas erschließen.



Abbildungen 34, 35: Schüler bei der Bearbeitung von „Archaeopteryx“.

e) Der Schaukasten „Archaeopteryx“ (vgl. Abbildung 34, 35):

Bezeichnung und Darstellung des „Schaufensters“ „Steinbruch“ suggerieren zunächst die idealisierte Nachbildung eines Steinbruchs im lithographischen Schiefer des Solnhofener Plattenkalks der Fränkischen Alb. Es werden typische Beispiele der Fauna präsentiert. Zu sehen sind Uhu (*Bubo bubo*) und Zauneidechse (*Lacerta agilis*). Dreidimensionale Strukturelemente (Schieferplatter, Gräser) im Vordergrund deuten ein Biotop an. Dennoch ist dieses „Schaufenster“ kein Diorama wie z.B. „Urwald im Fichtelgebirge“. Hintergrundbild und Legende der dargestellten Tiere fehlen. Nur der Uhu ist als „heimlicher Kulturfolger“ im Kontext des Biotops „Steinbruch“ behandelt. Die Gestaltung ist äußerst schlicht und in allen wesentlichen Punkten auf ein Thema ausgerichtet: Archaeopteryx. Die Gestaltungsform könnte man als Hybrid zwischen Diorama und Schaukasten charakterisieren. Die klare thematische Ausrichtung, die das Biotop zur Kulisse macht, veranlasst v. a. dazu, dieses Schaufenster als Schaukasten „Archaeopteryx“ zu bezeichnen. In den Schieferplatten ist der Abguss einer Versteinerung eines Archaeopteryx in originaler Größe zu sehen. Es handelt sich um einen Abguss des „Berliner Exemplars“ von 1877 (vgl. Bezzel & Prinzing 1990, S. 418; vgl. Hecht et al. 1985, S. 11). Im zentralen Erklärungstext des Schaukastens wird der Archaeopteryx als „missing link“, ein Bindeglied zwischen Vögeln und Reptilien, vorgestellt. Seine historische Bedeutung zur Stützung der Evolutionstheorie Darwins „im entscheidenden Moment“ wird referiert. Weiter heißt es: „Archaeopteryx vereinigt Reptilien- und Vogelmerkmale. Für Vögel typisch ist das Federkleid und die Form der Federn, das verwachsene Schlüsselbein, Reptilienmerkmale sind der lange Schwanz, die Fingerkrallen, Zähne und weitere Skelettmerkmale.“ Des Weiteren wird in drei Tabellen, die jeweils in räumlicher Nähe des Uhus als

Vogel, der Zauneidechse als Reptil und des Archaeopteryx als das „missing link“ angebracht sind, das Vorhandensein bzw. morphologische Besonderheiten folgender Punkte dargestellt: Schwanzwirbel, Bauchrippen, Rippen, Mittelfußknochen, Knochen (gemeint ist Pneumatisierung oder fehlende Pneumatisierung), Finger, Schlüsselbeine, Zähne, Federn und Kleinhirn. (Die identische Reihenfolge und ähnliche Wortwahl deuten auf Bezzel & Prinzing (1990, S. 419 f.) als Literaturquelle hin.) Besonders wichtige und für den Museumsbesucher einsichtige Merkmale wie das Vorhandensein von Federn oder Zähnen hätte man in der Reihung voranstellen können, um dadurch eine didaktisch sinnvolle Gewichtung der Merkmale zu bewirken. Die Auswahl der Merkmale erscheint vernünftig. Die Darstellung ist im Grunde fachlich akzeptabel, wenngleich kleinere „Unüblichkeiten“ zu bemerken sind: Bei „Vögel“ heißt es z.B. der „Mittelfußknochen“ sei „mit [dem] Wurzelfußknochen verwachsen.“ Die museale Darstellung folgt Prinzipien sektoraler und struktureller didaktischer Reduktion (vgl. Killermann 1995, S. 255 ff.). Im Museum behandelt man zehn Merkmale, Bezzel & Prinzing (1990, S. 419 f.) nennen 20 (sektorale didaktische Reduktion). Im Erklärungstext des Museums ist von „langem Schwanz“ der Reptilien die Rede und nicht von zahlreichen freien Schwanzwirbeln statt eines Pygostyls (vgl. Bezzel & Prinzing 1990, S. 417 und 419) (strukturelle didaktische Reduktion). Beide Maßnahmen erscheinen vernünftig und sind korrekt umgesetzt.

Etwas problematisch ist die unklare formale Struktur des „Schaufensters“, was didaktische Schwächen bedingt. Hätte man auf das Stopfpräparat des Uhus als Bewohner des Biotops Steinbruch verzichtet, und sich stattdessen für einen Vogel, der eher der Größe und Gestalt eines Archaeopteryx ähnelt, entschieden, z.B. eine Türkentaube oder eine Amsel, so würde eine Gegenüberstellung der Merkmale erheblich leichter fallen. Analoges gilt für die Wahl der Zauneidechse als Reptilienvertreter. Allein in etwa gleiche Größe der drei zu vergleichenden Exponate (Vogel, Reptil, Archaeopteryx) wäre für den Besucher hilfreich. Zusätzlich könnte man sich als methodische Unterstützung jeweils präparierte entsprechend der genannten Merkmale markierte Skelette von Vogel und Reptil vorstellen, um dem Besucher die Gegenüberstellung der Skelettmerkmale zu erleichtern. Die Abbildung einer Rekonstruktion eines Archaeopteryx, z.B. aus dem Jura-Museum in Eichstätt oder nach Bezzel & Prinzing (1990, S. 416 f. und 420: Skizzen von Rekonstruktionen und Skizze des Skeletts) würde dem Besucher des Naturkundemuseums des Lindenhofs weit größere An-

schaulichkeit bieten können. Fazit: Fachlich ist die Darstellung grundsätzlich korrekt, didaktisch-methodisch könnte man den Schaukasten optimieren.

Bei den limitierten Auswahlmöglichkeiten von didaktisch grundsätzlich akzeptablen Schaufenstern im Naturkundemuseum des Lindenhofs, zählt „Archaeopteryx“ noch zu den sechs am ehesten geeigneten. Trotz aller Schwächen wurde dieser Schaukasten in Voruntersuchungen von den Schülern sehr gut angenommen. Darum wird „Archaeopteryx“ auch in dieser Hauptuntersuchung als eine der Arbeitsstationen eingesetzt.



Abbildungen 36, 37: Schüler bei der Bearbeitung von „Platzbedarf“.

f) Der Schaukasten „Platzbedarf“ (vgl. Abbildung 36, 37):

In dieser Station – in der vorliegenden Untersuchung als „Schaukasten Platzbedarf“ bezeichnet – wird in der Hauptsache der Flächenbedarf von Tieren dargestellt. Daneben deutet man „die Arten-Areal-Beziehung“ an (Reichholf 1998, S. 101 ff.; vgl. Odum & Reichholf 1980, S. 51 ff.; Reichholf 1980, S. 13 ff.), die jedoch beim Museumsbesuch der Schüler keine Rolle spielt, und darum in dieser Beschreibung vernachlässigt werden kann. Bereits eine oberflächliche Plausibilitätsüberprüfung zeigt die Richtigkeit des Hauptinhalts des Schaufensters. „As one would expect, home range increases with body size“ (Peters 1983, S. 172, vgl. Brown 1995). Der Platzbedarf bzw. Flächenbedarf steigt mit zunehmender Körpergröße bzw. mit dem Gewicht eines Tieres. In der Darstellung bleibt man richtigerweise zur Verdeutlichung bei einer Tiergruppe, hier in der Klasse der Vögel. Bei Säugetieren ist analog derselbe Zusammenhang nachgewiesen (Peters 1983, S. 170 ff.). Im Schaukasten spricht man zunächst von Flächenbedarf, bezieht dann aber die Hauptgraphik und den Er-

klärungstext auf die Reviergröße. Es wird an keiner Stelle der Begriff „Revier“ (bzw. Territorium) definiert, z.B. als „gegen Artgenossen des gleichen Geschlechts verteidigtes Mindestwohngebiet eines Tieres“ (Schaefer 1992, vgl. Immelmann 1982). Sich bei der darzustellenden Hauptaussage auf das Territorium zu beziehen ist im Grunde korrekt, Heimbereich bzw. „home range“ als „normaler Aktivitätsbereich eines Individuums“ (Schaefer 1992) wäre eine andere ebenfalls fachlich richtige Möglichkeit gewesen (Peters 1983, 172 f.). Dennoch gibt es Unschärfen in der Museumsdarstellung: Beispielsweise vermengt man Vögel unterschiedlicher Ernährungstypen. „(...) carnivores have larger home ranges than herbivores of similar size“ (Peters 1983, S. 172). Dieser Zusammenhang gilt sicher nicht nur für „home ranges“, sondern ist auch für „Territorien“ von Karnivoren und Herbivoren zu erwarten. In der Darstellung des Museums hätte man entweder ähnliche Ernährungstypen darstellen sollen oder diese vielleicht aus Gründen falsch verstandener didaktischer Reduktion bewusst eingegangene Ungenauigkeit wenigstens im Erläuterungstext erwähnen sollen. Die logarithmische Skala der Ordinate findet sich in den Darstellungen von Brown (1995), als „territory or home range“, und in der von Peters (1983), als „home range“, ähnlich wie in der Hauptgraphik im Lindenhof. In der musealen Darstellung jedoch wird durch den Verlauf des den Zusammenhang beschreibenden Graphs ein „zusätzlicher“ exponentieller Anstieg der Territoriumsgröße trotz der bereits logarithmischen Skalierung der Ordinate angedeutet. Hier unterscheidet sich die Lindenhofdarstellung von den Angaben in der Literatur (Brown 1995, Peters 1983). Die Abszisse weist eine ungleichmäßige Skalierung auf, die eher den räumlichen Erfordernissen der Ausstellung der Stopfpräparate denn nachvollziehbaren mathematischen Überlegungen gehorcht. Die Präparate selbst befinden sich in einem auffällig guten Zustand. Dargestellt sind in der Reihe aufsteigender Körpergröße bzw. Revierbedarfs Blaumeise (*Parus caeruleus*), Grünfink (*Carduelis chloris*), Kernbeißer (*Coccothraustes coccothraustes*), Pirol (*Oriolus oriolus*), Grünspecht (*Picus viridis*) und Wespenbussard (*Pernis apivorus*). Überprüfungen der im Museum verwendeten Werte für Territoriumsgröße und Gewicht nach Bezzel (1985) und Wüst (1970) ergeben folgendes:

	Territorium [ha]		Gewicht [g]	
	Nach Bezzel	Lindenhof	Nach Wüst	Lindenhof
Blaumeise	0,5	<1	10-12	10
Grünfink	0,4-1,2	Ca. 1	25-35	25
Kernbeißer	0,5-5	1-10	50-57	50
Pirol	17	<10	68-78	70
Grünspecht	20-30	Ca. 100	185-210	150
Wespenbussard	Ca. 100	Ca. 1000	700-800	600

Hinsichtlich des Gewichts stimmen die Daten des Museums mit den recherchierten in etwa überein. Bei den aus der graphischen Darstellung im Museum nur sehr ungenau abzulesenden Territoriengrößen gibt es bisweilen erhebliche Unterschiede in den Angaben.

Bilanzierend lässt sich über den Schaukasten „Platzbedarf“ bezüglich der Grundaussage fachliche Richtigkeit attestieren. Bei Überprüfungen von Details und auf der Ebene einzelner Arten sind Schwächen nachweisbar. Die Stopfpräparate sind in besonders gutem Zustand. Für diese Untersuchung ist der Schaukasten wegen seiner klaren Hauptaussage besonders interessant.

1.2.4. Lernen ökologischer Inhalte

Diese Vorgaben des Naturkundemuseums erlauben die Entwicklung von Modulen, im Folgenden als Stationen bezeichnet, die in der punktuellen Intervention eines Museumsbesuchs das Lernen von Ökologie befördern. Präziser sollte man bei den o. g. Beispielen heimischer Lebensräume und ökologischer bzw. umwelterzieherisch relevanter Themen eher von Ausschnitten ökologischer Inhalte sprechen. Das Lernen ökologischer Inhalte, der Erwerb kognitiven Wissens, stellt eine, vielleicht die entscheidende Grundlage jeder Umwelterziehung dar. Umwelterziehung ist die Aufgabe, „Einstellungen und Verhalten der Menschen zu ändern“, um Probleme von Belastung und Zerstörung der Umwelt zu bewältigen (Killermann 1995, S. 122). Ziel ist v. a., die Handlungen von Menschen positiv zu beeinflussen. Von Becker (2000, S. 24) wird Umwelterziehung bzw. die Hinführung zu ökologischer Bildung in der Schule sehr kritisch gesehen: „Vor dem Hintergrund der (...) Sozialisationsfunktion der Schule,

(...) muss man sich die Frage stellen, ob es überhaupt in dieser traditionellen (...) staatlichen Institution möglich ist, ökologische Bildung zu vermitteln; ja ob es überhaupt sinnvoll ist, es zu versuchen (...). Weiter führt Becker (2000, S. 24) aus, „dass man sich keine Illusionen machen darf im Hinblick auf wirksame ökologische Lernprozesse, die über (kurzlebige) Wissensvermittlung zu bleibendem Bewusstsein und Verhaltensänderung führen.“ Nach Kaiser & Fuhrer (2000, S. 67 und S. 65) kommt dagegen „(...) der Wissensvermittlung im Rahmen von Umwelterziehung (...) höchste Priorität zu“, denn „Wissen [ist eine] (...) notwendige Bedingung ökologischen Handelns“ (vgl. Lantermann et al. 1992, zitiert nach Haase & Bogner 2002). Um umweltschützerische handlungsleitende Dispositionen zu entwickeln, sind in unterschiedlichen didaktischen Konzepten zur Umwelterziehung im zu Erziehenden verschiedene Zugänge zur Umweltwahrnehmung und zum Umweltverstehen zu eröffnen. Immer geht es um ein Zusammenwirken von unterschiedlichen Komponenten, bei Killermann (2000, S. 21 ff., 1995, S. 131) von Wissen, Erleben und Wertvorstellungen, als gleichrangige Kategorien dargestellt, bei Klautke & Köhler (1991, S. 48 ff.) von ähnlichen Komponenten, die im Unterschied zu Killermann als hierarchisches Stufenmodell angeordnet sind. Das Wissen spielt in beiden Konzepten eine entscheidende Rolle. Zur Vermittlung dieses Wissens schlägt Killermann (2000, S. 24 f.) im Rahmen eines „ganzheitlichen“ Konzeptes die „kognitionspsychologische Lerntheorie (...) Konstruktivismus“ als eine der „neuere[n] Formen des Lehrens (...) [vor], die gerade für die Umwelterziehung (...) wesentlich erschein[t]. (...) Eine zentrale Forderung der Konstruktivisten ist selbstgesteuertes Lernen statt darbietendem Unterweisen.“ Dieser Ansatz, insbesondere in einem außerschulischen Lernort verwirklicht, entspricht weitgehend den von Becker (2000, S. 13 ff.) geäußerten Vorstellungen, von Selbstbestimmung und außerschulischen Ansätzen zur Erreichung von Umweltbildung. Wissen allein ist jedoch nicht ausreichend, das eigentliche Ziel, umweltschützerisches Handeln, zu bewirken (Kaiser & Fuhrer, 2000, S. 67) bzw. bedingt es nur zu einem Teil (Lude 2001, S. 211 f., vgl. Haase & Bogner 2002, S. 83). Die Voraussetzung zur Organisation der Vermittlung dieses jedoch notwendigen kognitiven Zugangs bieten die sechs oben beschriebenen „Schaufenster“.

Die Ansprüche von Naturkundemuseum und Umwelterziehung treffen sich nicht nur in der Intention der Weitergabe kognitiven Wissens, insbesondere die affektive Ebene spielt für beide eine wesentliche Rolle. Museen haben als „Bereitsteller“ organi-

sierter authentischer Erfahrung die Aufgabe, Besucher in affektiver Hinsicht anzusprechen. Nach G. König (1999, S. 161) ist (bezogen auf Kinder- und Jugendmuseen) der Transport einer Botschaft besonders wichtig: „Es macht Spaß, ein Museum zu besuchen!“ Diese angestrebte positive Attribuierung darf nicht – quasi als Selbstzweck – „nur“ auf das Museum gerichtet bleiben. Entscheidend ist es, „sinnliche (...) Lernanlässe“ (Nuissl 1988, zitiert nach Gries 1996) auf die zu vermittelnden Inhalte zu richten. Die „affektive Komponente“ „Erleben“ ist nach Killermann (1995, S.130 f.) für die Umwelterziehung „von größter Bedeutung“, wie für Klautke & Köhler (1991) die affektiven Zugänge „Erleben“ und „Betroffensein“ als „Stufen“ zum Handeln unumgänglich sind. Für die Umwelterziehung stellen Museen damit eine sehr gute Ergänzung zu Aktivitäten im Freiland dar, auch wegen jahreszeiten- und witterungsunabhängiger Planbarkeit. Darin kann gerade für Exkursionen mit Schulklassen ein nicht unwesentlicher Vorteil bestehen. Für die vorliegende Untersuchung waren ebenfalls „diese Laborbedingungen in authentischer Situation“ von Vorteil.

2. Zum Unterricht

Die didaktische Operationalisierung der oben ausgeführten epistemologischen, pädagogischen und psychologischen Grundideen in Form unterrichtlicher Umsetzung als drei Treatments kann im Folgenden dreifach nachvollzogen werden, nämlich

- anhand der Arbeitspakete der Schüler,
- anhand einer Charakterisierung durch Prozessmerkmale des gemäßigten Konstruktivismus und
- anhand von Lernzielen, einem „traditionellen“ Vorgehen instruktional orientierten Unterrichts.

2.1. Die Aufgaben der Schüler

Die Arbeitspakete der Schüler lassen sich am konkretesten anhand der zu bearbeitenden Arbeitsblätter nachvollziehen. Laut Killermann (1995, S. 237 f.) wird der Einsatz von Arbeitsblättern nicht ganz unkritisch gesehen. Der Konflikt rankt sich v.a. um die Sorge zu großer Gängelung der Schüler und berührt damit einen Hauptansatz dieser Studie. Gries (1996, S. 12) schlägt vor, „Arbeitsblätter nach Vermittlungskonzepten zu gestalten, die von den Bedürfnissen und Fähigkeiten von Kindern und Ju-

gendlichen ausgehen und gleichzeitig museumsspezifische Angebote wie Begegnungen mit dem Original und authentische Erfahrung berücksichtigt.“ All diese Forderungen werden mit dem Konzept der gemäßigten Position des Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001) besonders im Auge behalten. Die Arbeitsblätter zu den drei Treatmentgruppen finden sich im Anhang 1. Als erstes werden die Vorlagen für das am ehesten instruktional ausgerichtete Vorgehen und das somit am wenigsten selbstbestimmte Treatment, das F-Treatment (F=Fremdbestimmung), dargestellt. Es folgt die Mischform, das SF-Treatment (SF=Selbst- und Fremdbestimmung), zuletzt das Treatment mit dem höchsten Maß an Selbstbestimmung, das am meisten konstruktivistischen Vorstellungen entspricht, das S-Treatment (S=Selbstbestimmung).

Anhand dieser Arbeitsblätter lassen sich die oben angesprochenen affektiven Zugänge durch „Erleben“ (Killermann 1995, Klautke & Köhler 1991) oder „Betroffenheit“ (Klautke & Köhler 1991) erahnen. Beispielsweise reagieren Kinder auf den narrativen Einstieg zur Station „Sandgrube“ z. T. mit Antworten in der Ersten Person, Singular (Ich-Form), die Figuren aus der Einstiegsgeschichte werden direkt angesprochen, z.B.: „Schauen sie mal die vielen Vögel und Insekten, die es hier gibt, [an]. Ich finde das einfach wunderbar. Die Tiere fühlen sich hier wohl.“ Dies zeigt, wie intensiv die vermeintlich rein kognitive Bearbeitung der Arbeitspakete von den Kindern durch „inneres Erleben“ begleitet werden kann und in welchem Maße die Probanden affektiv angesprochen werden können.

2.2. Treatmentcharakterisierung nach Kriterien des gemäßigten Konstruktivismus

Bezogen auf die von Reinmann-Rothmeier & Mandl (2001, S. 626) formulierten „Prozessmerkmale“ einer gemäßigt konstruktivistischen Position (Selbststeuerung, Eigenaktivität, Situiertheit (unter Einbeziehung der sozialen Dimension), Konstruktion eigener Wirklichkeit) werden die Treatments analysiert.

Das F-Treatment (v. a. fremdbestimmt) beginnt wie auch die beiden übrigen Treatments mit einer einführenden Geschichte von wenigen Zeilen. In diesem Treatment bearbeiten die Schüler in einer Arbeitsstation bis zu acht Multiple-Choice-Aufgaben

mit bis zu zwölf Antwortmöglichkeiten, wobei die Anzahl richtiger Antworten unbekannt ist. Global formuliert ist das F-Treatment gekennzeichnet durch den geringsten Grad an *Selbststeuerung*. Das bedeutet, eine weitgehende Fremdbestimmung der einzelnen Handlungsabschnitte mit geringen Wahlmöglichkeiten für die Probanden. Sie haben weder Einfluss auf das Ziel noch auf die Organisation der Einzelschritte des Lernprozesses. Auch die externe Koordination, gemeint ist die organisatorische Einbettung in weitere Aufgaben (vgl. Reinmann-Rothmeier & Mandl 2001, S. 634), ist fremdbestimmt. Einzig über den eigentlichen Prozess der Wissensaufnahme könnten die Schüler entscheiden, indem sie die Aufgaben bewusst „ohne Verstand“ durchführten. Die weitgehende Fremdbestimmung wirkt sich auf den Grad der *Eigenaktivität* aus. Den Schülern ist lediglich abverlangt, richtige Lösungen zu kennzeichnen, z.B. durch Ankreuzen. Inwieweit besonderes situatives Interesse geweckt wird, ist nicht zu beurteilen. Da jedoch der museale Rahmen und auch die Einstiegsgeschichte, vgl. Anchored Instruction (Reinmann-Rothmeier & Mandl 2001, S. 617 f.), zu jeder Station bei den drei Treatments identisch ist, sollte das situative Interesse für die drei Treatments vergleichbar hoch sein. *Situiertheit* bedingt durch Authentizität des Kontextes ist durch die Gestaltung der Dioramen weitgehend gegeben. Heimische Biotope werden in idealisierter Form nachgebildet. Originale Repräsentationsformen werden im nachgeahmten Wirklichkeitszusammenhang gezeigt (vgl. Uhlig 1962, S. 68). Die Schaukästen weisen einen weniger hohen Grad an Authentizität auf. Neben erscheinungsaffinen Formen wie Farbfotos werden hier auch inaffine Formen wie Diagramme verwendet (Nomenklatur nach Uhlig 1962, S. 69 ff.). Die Sozialform ist nicht vorgegeben, wird aber sicher durch die Gruppengröße von drei oder vier Schülern und die Art der Aufgaben beeinflusst. Die Anforderungen bereiten den Schülern i.d.R. inhaltlich keine Probleme. Sie verstehen ohne Schwierigkeiten, was von ihnen verlangt wird und sind den Anforderungen gewachsen, so sind sie nicht aufgrund von Missverständlichkeiten oder zu schwerer Aufgaben veranlasst – gewissermaßen einem latenten Druck gehorchend - mit den anderen Schülern der Gruppe zusammenzuarbeiten. Der soziokulturelle Kontext der Schüler sollte für die drei Treatmentgruppen in identischer Weise bedeutsam sein, so dass dieser Einflussfaktor im Weiteren nicht mehr behandelt werden wird. Sicher wird durch die eng geführten Aufgaben mit wenigen Möglichkeiten, selbst etwas zu bewirken, die *Konstruktion eigener Wirklichkeit* beeinflusst.

Im SF-Treatment (Mischform zwischen Selbst- und Fremdbestimmung) werden den Probanden pro Station bis zu fünf Aufgaben gestellt, die z.B. den Vergleich zweier Tiere behandeln oder die Ergänzung von Skizzen fordern. Ein mittlerer Grad der Selbstbestimmung ist verwirklicht. Jede Aufgabe gibt den Probanden zunächst ein fremdbestimmtes Korsett vor, das grobe Ziel ist festgesetzt, jedoch in der Ausgestaltung der Bearbeitung, in der Organisation des Lernprozesses und in der Festlegung der Einzelergebnisse haben die Lerner die Möglichkeit, selbstbestimmt zu entscheiden. Die *Eigenaktivität* kann sich im Schreiben von Texten, im Sinne von Lienert (1969, S. 28) von „Kurzaufsätzen“, und Zeichnen jeweils im Rahmen der von der Aufgabe vorgegebenen Grenzen ausdrücken. Wie beim F-Treatment ist bezüglich des objektbezogenen Kontextes *Situietheit* gegeben. Die soziale Ebene betreffend ist zwar das äußere Bedingungsgefüge vergleichbar, könnte aber dennoch durch die sich unterscheidende Aufgabenstellung unterschiedliche „Kooperationsbedürftigkeit“ bedingen. Eine *Konstruktion eigener Wirklichkeiten* könnte durch den Freiraum gefördert werden, könnte aber stattdessen auch – wie möglicherweise im F-Treatment – durch das Korsett der Aufgaben eingengt sein.

Das höchste Maß an Selbststeuerung / -bestimmung bietet das S-Treatment. An die kurze narrative Hinführung anknüpfend oder neu ansetzend wird den Schülern eine möglichst allgemein gehaltene Aufgabe gestellt, wie z.B. „(...) Schreibe auf, was Dich besonders interessiert (...)!“ oder „Verstehst Du das Schaufenster? (...) Schreibe möglichst genau auf, was (...) dargestellt sein soll!“ Damit wird den Schülern ein hoher Grad an Autonomie gestattet. Der Lerner wählt den zu lernenden konkreten Inhalt weitgehend selbst aus, hat damit entscheidenden Einfluss auf das Ziel des Lernprozesses. Im Rahmen der zeitlichen Vorgaben und der gegebenen Gruppengröße von drei bis vier Schülern kann der Proband seinen Lernprozess koordinieren und organisieren, wie es ihm beliebt. Einzig Überlegungen der Schüler auf Meta-Ebenen zu Bearbeitung und Möglichkeiten einer autonomen Lernerfolgskontrolle werden nicht explizit angestoßen. Aus diesen Rahmenvorgaben kann sich ein hohes Maß an *Eigenaktivität* entwickeln, da die Ausgestaltung der Bearbeitung den Schülern jeden Freiraum lässt. Sie müssen sich zunächst die Art der Bearbeitung überlegen: Texte, Zeichnungen, Tabellen, Diagramme – all dies ist möglich und wird auch gezeigt. Insbesondere die direkte Aufforderung, den Gegenstand des eigenen Interesses bzw. des persönlichen Verstehens zu beschreiben, macht es für eine sinnvolle Bearbei-

tung der Aufgabe unumgänglich, sich mit den Inhalten aktiv auseinander zu setzen. Diese Situation könnte mehr Austausch zwischen den Schülern erfordern als die zuvor beschriebenen Treatments. Ein Aspekt der *Situiertheit* weicht möglicherweise deutlich von den anderen Treatments ab. Die persönliche Involviertheit könnte die *Konstruktion der Wirklichkeit* der Individuen positiv beeinflussen.

2.3. Treatmentcharakterisierung nach Kriterien instruktional orientierten Unterrichts

2.3.1. Vorbemerkung

Hierbei geht es um die Formulierung und die Erfüllbarkeit von Lernzielen. „Unter Lernziel versteht man (...) [die] Beschreibung von Eigenschaften, die der Lernende nach erfolgreicher Lernerfahrung erworben hat“ (Mager 1965, S. 3). „Ein Lernziel legt (...) fest, was der Schüler wissen und können soll, welche Einsichten er gewinnen und welches Verhalten er zeigen soll“ (Killermann 1995, S. 32). Die Grundidee dieser Definition hat sich in den letzten Jahrzehnten offensichtlich kaum verändert.

Im Folgenden werden Lernziele von zwei Abstraktionsebenen formuliert, um die erreichten kognitiven Leistungen der Schüler zu präzisieren: Grobziele und Feinziele. Zwei hierarchisch übergeordnete Ebenen, Leitziele und Richtziele (Westphalen 1980, S. 53), werden aufgrund der hier fehlenden unmittelbaren Relevanz nicht präzise entwickelt. Ein Leitziel könnte z.B. sein: „Die Schüler sollen umweltbewusste Bürger werden!“ Ein Richtziel, z.B.: „Die Schüler sollen typische Vertreter der heimischen Fauna (...) kennenlernen“ (Killermann 1995, S. 32). Auf der folgenden Ebene stehen die Grobziele. Diese sind, wie Leit – und Richtziele, übergeordnete und nicht operationalisierbare Lernziele. Aus ihnen lassen sich Feinziele entwickeln. Feinziele sind detaillierte Ziele des Unterrichts, die überprüft werden können (Definitionen vgl. Berck 2003). Diese Überprüfbarkeit bzw. Operationalisierbarkeit der Feinziele gilt als entscheidendes Kriterium für Unterrichtseffekte. Für die Grobziele werden Zielklassen und Anforderungsstufen gemäß der Nomenklatur und der Matrix von Westphalen (1980, S. 47) verwendet. In den Grobzielen findet ausschließlich die Zielklasse „Wissen“ des kognitiven Bereiches der oben angesprochenen Grobzielmatrix Berücksichtigung (Begründung siehe folgender Absatz). Mit Bedacht wird in allen Fällen eine

nicht zu hohe „Anforderungsstufe“ verwendet. Der Wissensstand der unbekannten Schüler ist – trotz der Vorgaben des Lehrplans - nicht präzise einzuschätzen. Außerdem ist der unterrichtliche Eingriff pro Diorama bzw. Schaukasten sehr kurz. Aus diesen Gründen erscheint es ratsam, die angestrebten Ziele nicht zu hoch anzusetzen. Die Feinziele werden in operationalisierbarer Form dargestellt.

Auf affektiver Ebene werden Lernziele nur auf der Grobzielebene formuliert (vgl. NSTA 1969). Der in Eschenhagen et al. (2001, S. 180 ff.) geschilderte Effekt einer negativen Beeinflussung affektiver Lernziele durch den Zwang, diese operationalisierbar zu formulieren, wird damit vermieden.

Weitere Kategorisierungen und Differenzierungen von Lernzielen bezüglich der Treatments, z.B. nach Bloom (Anderson et al. 2001), auch psychomotorische Lernziele (nach Westphalen 1980, S. 51 f.), sollen bei der Lernzielformulierung keine Berücksichtigung finden. Die in diesem Museum zu verwirklichenden psychomotorischen Lernziele beziehen sich auf keine im Besonderen oder ausschließlich dem Fach Biologie zuzuordnenden fachgemäßen Arbeitsweisen (im Sinne von Killermann 1995, S. 196 ff.). Darum wird auf deren Formulierung verzichtet.

Zu den Grobzielen ist noch ein weiterer Aspekt anzusprechen. Es lassen sich in dieser Untersuchung zwei Ebenen von Grobzielen unterscheiden: explizite und implizite. Mit expliziten Grobzielen, sind Lernziele gemeint, die durch die im Treatment gestellten Aufgaben absichtsvoll erreicht werden sollen. Sie lassen sich in Form von Feinzielen klar operationalisieren. Mit impliziten Grobzielen sind Lernziele gemeint, die innerhalb der organisatorischen und inhaltlichen musealen Vorgaben durchaus erreicht werden können, aber keine explizite Manifestation in den instruktionalen Vorgaben des Treatments haben. Es handelt sich hier nicht primär um unterschiedliche Stufen kognitiver Anforderung, die bekannter Hierarchisierung kognitiver Lernziele folgen, z.B. der Stufenfolge Reproduktion, Reorganisation, Transfer und Problemlösen (vgl. Eschenhagen et al. (2001, S. 180), sondern es geht eher darum, ob Schüler neben dem eigentlichen Fokus ihrer Aufgabe andere Inhalte „mit-aufnehmen“ können.

2.3.2. Die Lernziele (Grobziele, Feinziele) der Treatments

Für die Darstellung ist es erforderlich, für jedes Diorama kurz das Aufgabenpaket der Schüler zu umreißen. Dies geschieht jeweils nach der Nennung des für alle Treatments geltenden Grobziels „*Grobziel_{kog}*“. Für manche Treatments entspricht es einem impliziten Lernziel. Erst dann folgt die Präzisierung der speziell für die jeweiligen Treatments geltenden Grobziele und die Formulierung auf der Ebene der Feinziele. Die folgende (recht umfangreiche) Charakterisierung wird durch einen etwas kleineren Schrifttyp angezeigt.

Im Folgenden wird v. a. bei den „Multiple-Choice-Aufgaben“ von „Distraktoren“ die Rede sein. Gemeint sind der korrekten Lösung ähnliche, aber unrichtige Alternativlösungen (vgl. z.B. Bortz & Döring 1995, S. 195 f.). Eine präzise Distraktorenanalyse wird nicht durchgeführt, denn es geht hier um die Treatments, die das Ergebnis der Untersuchung „nur“ mittelbar bestimmen, und nicht um die Fragebögen, die unmittelbar Einfluss auf die Testresultate haben.

Urwald im Fichtelgebirge

Grobziel_{kog}: Einblick in den Lebensraum „Urwald im Fichtelgebirge“

Grobziel_{aff}: Achtung vor dem Lebensraum „Urwald im Fichtelgebirge“

Treatment F:

Narrativer Einstieg

- Finde (...) den Schwarzspecht!
- Welcher der abgebildeten Vögel gehört zu den Spechten? (...) – Silhouette / Habitus von Schwarzspecht und zwei Distraktoren
- Welcher Fuß gehört einem Specht? (...) – Schwarzweißbild eines Schwarzspechtfußes und zwei Distraktoren
- Welcher Schnabel gehört einem Specht? (...) - Schwarzweißbild eines Schwarzspechtschnabels und zwei Distraktoren
- „Frankensteinspiel“

Grobziel_F: Grobziel_{kog} latentes Lernziel; explizites Grobziel_F: Kenntnis von Merkmalen eines Schwarzspechts

Feinziele:

Die Schüler sollen...

- die Artbezeichnung eines Spechtes nennen können.
- das Stopfpräparat eines Schwarzspechts seiner Artbezeichnung zuordnen können.
- den Habitus eines Schwarzspechtes anhand von Schwarzweißabbildungen neben zwei Distraktoren herausfinden und anzeigen können.
- eine Schwarzweißabbildung eines Schwarzspechtfußes neben zwei Distraktoren herausfinden und anzeigen können.
- eine Schwarzweißabbildung eines Schwarzspechtschnabels neben zwei Distraktoren herausfinden und anzeigen können.

Treatment SF:

Narrativer Einstieg

- Finde (...) den Buntspecht! (...)
- Wie viele Spechte sind (...) zu sehen?
- Welche Spechte sind (...) zu sehen?
- Wie unterscheiden sich Buchfink und Buntspecht? (...)
- Korrigiere das Vogelbild (...) (zu einem Specht)! – Kopf und Rumpf sind angedeutet.

Grobziel_{SF}: Grobziel_{kog} latentes Lernziel; explizites Grobziel_{SF}: Einblick in die Tiergruppe (Arbeitspiele, Merkmale) der Spechte

Feinziele:

Die Schüler sollen...

- die Artbezeichnung von drei Spechten nennen können.
- die Stopfpräparate von Buntspecht, Schwarzspecht und Grünspecht den Artbezeichnungen zuordnen können.
- drei Spechtmerkmale nennen können.
- drei Spechtmerkmale zeichnen können.

Treatment S:

Narrativer Einstieg

- Erkläre dem Indianer den „Urwald im Fichtelgebirge“! Schreibe einfach auf, was Dich besonders interessiert, was vielleicht auch für Dich neu ist. Schau es Dir genau an und mache Dir Notizen. Was willst Du ihm erzählen?

Grobziel_S entspricht Grobziel_{kog} (Einblick in den Lebensraum „Urwald im Fichtelgebirge“)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- Inhalte zum Urwald im Fichtelgebirge zum Ausdruck bringen können.

Wacholderheide

Grobziel_{kog}: Einblick in den Lebensraum „Wacholderheide im Frankenjura“

Grobziel_{aff}: Achtung vor dem Lebensraum „Wacholderheide im Frankenjura“

Treatment F:

Narrativer Einstieg

- (...) Finde (...) den Raubwürger!
- Welcher der abgebildeten Vögel ist der Würger? (...) – Silhouette / Habitus von Raubwürger und zwei Distraktoren
- Welcher Schnabel gehört dem Würger? (...) - Schwarzweißbild eines Raubwürgerschnabels und zwei Distraktoren
- Welche Kopfzeichnung hat der Würger? (...) - Schwarzweißbild eines Raubwürgerkopfes und zwei Distraktoren
- „Frankenstein Spiel“

Grobziel_F: Grobziel_{kog} latentes Lernziel; explizites Grobziel_F: Kenntnis von Merkmalen eines Raubwürgers

Feinziele:

Die Schüler sollen...

- die Artbezeichnung eines Würgers nennen können.
- das Stopfpräparat eines Raubwürgers seiner Artbezeichnung zuordnen können.
- den Habitus eines Raubwürgers anhand von Schwarzweißabbildungen neben zwei Distraktoren herausfinden und anzeigen können.
- eine Schwarzweißabbildung eines Raubwürgerschnabels neben zwei Distraktoren herausfinden und anzeigen können.
- eine Schwarzweißabbildung der Kopfzeichnung eines Raubwürgers neben zwei Distraktoren herausfinden und anzeigen können.

Treatment SF:

Narrativer Einstieg

- (...) Finde (...) den Neuntöter! (...)
- Wie viele Würger sind (...) zu sehen?
- Schreibe die Namen der Würger auf (...)!
- Vergleiche Steinschmätzer und Neuntöter! (...)
- Korrigiere das Vogelbild (...)! (...) – Kopf und Rumpf sind angedeutet.

Grobziel_{SF}: Grobziel_{kog} latentes Lernziel; explizites Grobziel_{SF}: Einblick in die Tiergruppe Würger

Die Schüler sollen...

- die Artbezeichnung von zwei Würgern nennen können.
- die Stopfpräparate von Neuntöter und Raubwürger der jeweiligen Artbezeichnung zuordnen können.
- drei Würgermerkmale nennen können.
- drei Würgermerkmale zeichnen können.

Treatment S:

Narrativer Einstieg

- Kannst Du Blaubär etwas zur „Wacholderheide im Frankenjura“ erzählen? Schreibe auf, was auch Dich besonders interessieren würde!

Grobziel_S entspricht Grobziel_{kog} (Einblick in den Lebensraum „Wacholderheide im Frankenjura“)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- Inhalte zur „Wacholderheide im Frankenjura“ zum Ausdruck bringen können.

Sandgrube

Grobziel_{kog}: Einblick in den Lebensraum „Sandgrube“

Grobziel_{aff}: Interesse am Lebensraum „Sandgrube“

Treatment F:

Narrativer Einstieg

- (...) Finde (...) den Eisvogel!
- Welche Gefiederzeichnung gehört dem Eisvogel? (...) - Schwarzweißbild eines Eisvogelkopfes und zwei Distraktoren
- Welche Körpergestalt hat der Eisvogel? (...) - Silhouette / Habitus des Eisvogels und zwei Distraktoren
- Welcher Schnabel gehört dem Eisvogel? (...) - Schwarzweißbild eines Eisvogelschnabels und zwei Distraktoren
- „Frankensteinispiel“

Grobziel_F: Grobziel_{kog} latentes Lernziel; explizites Grobziel_F: Kenntnis von Merkmalen eines Eisvogels

Feinziele:

Die Schüler sollen...

- das Stopfpräparat eines Eisvogels seiner Artbezeichnung zuordnen können.
- eine Schwarzweißabbildung der Gefiederzeichnung eines Eisvogels neben zwei Distraktoren herausfinden und anzeigen können
- den Habitus eines Eisvogels anhand von Schwarzweißabbildung neben zwei Distraktoren herausfinden und anzeigen können.
- eine Schwarzweißabbildung eines Eisvogelschnabels neben zwei Distraktoren herausfinden und anzeigen können

Treatment SF:

Narrativer Einstieg

- (...) Finde (...) den Eisvogel! (...)
- Wie viele Eisvögel findest Du?
- Beschreibe Unterschiede zwischen Eisvogel und Bienenfresser! (...)
- Korrigiere das Vogelbild (...)! (...) - Kopf und Rumpf sind angedeutet.

Grobziel_{SF}: Grobziel_{kog} latentes Lernziel; explizites Grobziel_{SF}: Kenntnis des Eisvogels

Feinziele:

Die Schüler sollen...

- das Stopfpräparat eines Eisvogels seiner Artbezeichnung zuordnen können.
- drei Eisvogelmerkmale nennen können.
- drei Eisvogelmerkmale zeichnen können.

Treatment S:

Narrativer Einstieg

(...) – Weiterführung der Geschichte bis zum Anlass für die Aufgabe

- Beschreibe (...)! Schreibe auf, was für Dich besonders Interessant ist, was vielleicht sogar für Dich neu ist. Schau das Schaufenster genau an und mache Dir Notizen. (...)

Grobziel_S entspricht Grobziel_{kog} (Einblick in den Lebensraum „Sandgrube“)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- Inhalte zur „Sandgrube“ zum Ausdruck bringen können.

Kulturfolger

Grobziel_{kog1}: Einblick in die Problematik der rückläufigen Bestände ehemaliger Kulturfolger in der heutigen mitteleuropäischen Kulturlandschaft

Grobziel_{kog2}: Einblick in Maßnahmen bzw. Faktoren, die eine „ökologisch intakte mitteleuropäische Kulturlandschaft“ begünstigen können

Grobziel_{aff}: Offenheit, sich mit den Fragen zu Kulturfolgern auseinanderzusetzen

Treatment F:

Narrativer Einstieg

- Welches Hasentier siehst Du im Schaufenster? (...)
- Finde (...) den Namen (...) heraus!
- (Hier erfolgt der Hinweis auf bestimmte Bilder im Schaukasten, nämlich ein Saumbiotop und eine Monokultur. Sie werden als Anlass und Bearbeitungshilfe für die folgende Aufgabe verstanden.) Diese „Landschaftselemente“ verbessern die Überlebenschancen von „Hasentieren“: Streiche Falsches durch! (Es folgt die Nennung von insgesamt 18 fördernder, kontraproduktiver oder neutraler Elemente.)

Grobziel_F: Grobziel_{kog1} latentes Lernziel, Grobziel_{kog2} entspricht Grobziel_F (Einblick in Maßnahmen bzw. Faktoren, die eine „ökologisch intakte mitteleuropäische Kulturlandschaft“ begünstigen können)

Feinziele:

Die Schüler sollen...

- ein Beispiel für einen typischen Kulturfolger nennen können.
- das Stopfpräparat eines Feldhasen seiner Artbezeichnung zuordnen können.
- anhand einer Liste von Begriffen die eine „ökologisch intakte mitteleuropäische Kulturlandschaft“ begünstigende und schädigende Faktoren herausfinden und anzeigen können.

Treatment SF:

Narrativer Einstieg

- Welche Tiere siehst Du im Schaufenster? Finde die Namen der Tiere heraus (...)!
 - (... → Foto einer Monokultur) Welche Schwierigkeiten könnten die Tiere auf diesem Feld haben?
 - (... → Schwarzweißabbildung einer flurbereinigten Landschaft) Verändere das Bild so, dass sich Tiere dort wohl fühlen. (...)

Grobziel_{SF}: Grobziel_{kog1} latentes Lernziel, Grobziel_{kog2} entspricht Grobziel_{SF} (Einblick in Maßnahmen bzw. Faktoren, die eine „ökologisch intakte mitteleuropäische Kulturlandschaft“ begünstigen können); (und: Einblick in die typische Fauna einer landwirtschaftlich geprägten mitteleuropäischen Kulturlandschaft)

Feinziele:

Die Schüler sollen...

- drei Tiere einer typischen mitteleuropäischen Kulturlandschaft nennen können
- drei Tiere (Goldammer, Rebhuhn, Feldhase) anhand von Stopfpräparaten ihren Artbezeichnungen zuordnen können.
- zwei „Schwierigkeiten“ von Tieren, die in Monokulturen leben müssen, nennen können.
- Strukturelemente einer „ökologisch intakten mitteleuropäischen Kulturlandschaft“ in die Abbildung einer durch Monokulturen geprägten Kulturlandschaft einzeichnen können.

Treatment S:

Narrativer Einstieg

- Kannst Du (...) erklären, wie man unsere Kulturlandschaft verändern müsste, damit Hasen und Rebhühner und viele andere Tiere hier wieder besser leben können? Schreibe Deine Erklärung auf!

Grobziels: Grobziel_{kog1} latentes Lernziel, Grobziel_{kog2} entspricht Grobziel_S (Einblick in Maßnahmen bzw. Faktoren, die eine „ökologisch intakte mitteleuropäische Kulturlandschaft“ begünstigen können)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- Ökologisch sinnvolle Verbesserungen für die Überlebensmöglichkeiten von Kulturfolgern bezüglich der Ist-Situation (so wie im Schaukasten dargestellt) in der mitteleuropäischen Kulturlandschaft vorschlagen können.

Archaeopteryx

Grobziel_{kog}: Einblick in Merkmale, die für eine phylogenetische Verwandtschaft von Archaeopteryx und Vogel bzw. Reptil sprechen

Grobziel_{aff}: Interesse an einer Auseinandersetzung mit dem Urvogel Archaeopteryx

Treatment F:

Narrativer Einstieg

(...) – Vorgabe von Information (Verwandtschaft zwischen Reptil und Archaeopteryx bzw. Vogel und Archaeopteryx)

- Streiche (...) die falsche Tiergruppe durch!
- Die Zähne des Archaeopteryx weisen auf eine Verwandtschaft mit Vogel / Reptil hin!
- Die Federn des Archaeopteryx weisen auf eine Verwandtschaft mit Vogel / Reptil hin!

- Die Krallen an den Flügeln des Archaeopteryx weisen auf eine Verwandtschaft mit Vogel / Reptil hin!
- (...) Kreuze an, welche von den Merkmalen der Archaeopteryx hatte. (...) – Federn, Zähne, Krallen an den Flügeln, „kurzes Schwänzchen“

Grobziel_F entspricht Grobziel_{kog} (Einblick in Merkmale, die für eine phylogenetische Verwandtschaft von Archaeopteryx und Vogel bzw. Reptil sprechen)

Feinziele:

Die Schüler sollen...

- bei Vorgabe des Merkmals „Zähne“ entscheiden können, ob dieses Merkmal des Archaeopteryx eher für eine phylogenetische Verwandtschaft mit einem Vogel oder einem Reptil spricht.
- bei Vorgabe des Merkmals „Federn“ entscheiden können, ob dieses Merkmal des Archaeopteryx eher für eine phylogenetische Verwandtschaft mit einem Vogel oder einem Reptil spricht.
- bei Vorgabe des Merkmals „Krallen an den Flügeln“ entscheiden können, ob dieses Merkmal des Archaeopteryx eher für eine phylogenetische Verwandtschaft mit einem Vogel oder einem Reptil spricht.

Treatment SF:

Narrativer Einstieg

(...) – Vorgabe von Information (Verwandtschaft zwischen Reptil und Archaeopteryx und Vogel und Archaeopteryx)

- Schreibe auf, welche Merkmale der Archaeopteryx mit dem Federvieh (...) und welche mit dem Echsentier (...) gemeinsam hat.
- Der Archaeopteryx ist ein

Grobziel_{SF} entspricht Grobziel_{kog} (Einblick in Merkmale, die für eine phylogenetische Verwandtschaft von Archaeopteryx und Vogel bzw. Reptil sprechen)

Feinziele:

Die Schüler sollen...

- drei Merkmale des Archaeopteryx nennen können.
- bei den selbst genannten drei Merkmalen jeweils entscheiden können, ob sie eher für eine Verwandtschaft mit einem Vogel oder mit einem Reptil sprechen.

Treatment S:

Narrativer Einstieg

- (...) – Vorgabe von Information (Verwandtschaft zwischen Reptil und Archaeopteryx und Vogel und Archaeopteryx)
- Kannst Du die beiden (Reptil / Vogel – ergänzt) überzeugen? Schreibe Deine Argumente auf!

Grobziel_S entspricht Grobziel_{kog} (Einblick in Merkmale, die für eine phylogenetische Verwandtschaft von Archaeopteryx und Vogel bzw. Reptil sprechen)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- Hinweise für eine phylogenetische Verwandtschaft des Urvogels Archaeopteryx mit einem Vogel bzw. mit einem Reptil nennen können.

Platzbedarf

Grobziel_{kog}: Kenntnis des Zusammenhangs zwischen Körpergröße und Platzbedarf⁸

Grobziel_{aff}: Bereitschaft, sich mit der Problematik des Raumbedarfs für Tiere auseinanderzusetzen

Treatment F:

Narrativer Einstieg

- (...) Wie viel Platz braucht ein Wespenbussard? – richtige Angabe und zwei Distraktoren
- Wie schwer ist ein Kernbeißer? – richtige Angabe und zwei Distraktoren
- Ein Spatz ist so groß wie ein Grünfink. Also wiegt er – richtige Angabe und zwei Distraktoren
- Ein Spatz braucht wahrscheinlich eine Reviergröße von – richtige Angabe und zwei Distraktoren
- Wie viel Platz braucht ein Grünfink? – richtige Angabe und zwei Distraktoren
- Wie schwer ist ein Wespenbussard? – richtige Angabe und zwei Distraktoren
- Wie viel Platz braucht ein Kernbeißer? – richtige Angabe und zwei Distraktoren
- Wie schwer ist ein Grünfink? – richtige Angabe und zwei Distraktoren

Grobziel_F entspricht Grobziel_{kog} (Kenntnis des Zusammenhangs zwischen Körpergröße und Platzbedarf)

Feinziele:

Die Schüler sollen...

- anhand einer Graphik das richtige Gewicht eines in der Graphik behandelten Vogels neben zwei Distraktoren herausfinden und anzeigen können.
- anhand einer Graphik den angemessenen Platzbedarf eines in der Graphik behandelten Vogels neben zwei Distraktoren herausfinden und anzeigen können.

Treatment SF:

Narrativer Einstieg

- Vergleiche Platzbedarf und Gewicht von Pirol und Wespenbussard! (...)
- Ein Zeisig ist ungefähr so groß wie eine Blaumeise. Wie schwer ist er wohl? Wie viel Platz braucht er ungefähr?
- (...) Schätze doch einmal, wie schwer eine Taube ist! Wie viel Platz braucht sie?

Grobziel_{SF} entspricht Grobziel_{kog} (Kenntnis des Zusammenhangs zwischen Körpergröße und Platzbedarf)

Feinziele:

Die Schüler sollen...

- anhand einer Graphik das richtige Gewicht eines in der Graphik behandelten Vogels bestimmen und nennen können.
- Gemeinsamkeiten bzw. Unterschiede im Gewicht zwischen zwei wie oben behandelten Vögeln nennen können.
- anhand einer Graphik den angemessenen Platzbedarf eines in der Graphik behandelten Vogels bestimmen und nennen können
- Gemeinsamkeiten bzw. Unterschiede im Platzbedarf zwischen zwei wie oben behandelten Vögeln nennen können.
- anhand einer Graphik das richtige Gewicht eines den Schülern bekannten, aber nicht in der Graphik behandelten Vogels bestimmen und nennen können.
- anhand einer Graphik den angemessenen Platzbedarf eines den Schülern bekannten, aber nicht in der Graphik behandelten Vogels bestimmen und nennen können

⁸ Der Terminus „Platzbedarf“ ist selbstverständlich jeweils im Kontext des entsprechenden Schaukastens zu verstehen. Im Schaukasten geht es um Territorien.

Treatment S:

Narrativer Einstieg

- Verstehst Du das Schaufenster? (...) Kannst Du (...) den Inhalt des Schaufensters erklären? Schreibe möglichst genau auf, was im Schaufenster dargestellt sein soll!

Grobziel_S entspricht Grobziel_{kog} (Kenntnis des Zusammenhangs zwischen Körpergröße und Platzbedarf)

Feinziele sind bei diesem Treatment nicht sinnvoll zu formulieren (vgl. Unterrichtliche Umsetzung 2.3.4.). Eine ungefähre operationalisierbare Zielvorgabe kann wie folgt lauten:

Die Schüler sollen...

- den Zusammenhang zwischen Platzbedarf und Körpergröße bei Vögeln schildern können.

2.3.3. Formulierung, Sprache und Didaktische Reduktion

In Sprache und Formulierung geht es nicht in erster Linie um semantische Perfektion, sondern um Verständlichkeit, z. T. um eine positive Attribuierung ohne Bedeutungs- bzw. Verständlichkeitsverlust für die Schüler. Hierzu seien zwei Beispiele angeführt und kurz analysiert.

1. Beispiel: Aufgabe des Treatment S / Urwald im Fichtelgebirge: „Erkläre (...) den ‚Urwald im Fichtelgebirge‘!“ –Sollten Schüler etwas wie ein theoretisch durchdachtes gedankliches Konstrukt von „Urwald“ besitzen, ist die Formulierung treffend gewählt. Das ist jedoch kaum zu erwarten. Die viel eher die wahrscheinliche Aktion des Schülers treffende Vokabel wäre vielleicht „Beschreibe (...)!“, „Berichte (...)!“, oder „Schildere (...)!“ Obwohl „erklären“ eher zur Kommunikation gedanklicher Konstrukte Verwendung findet, wertet es hier die wahrscheinliche Beschreibung des Schülers ohne Bedeutungsverlust auf und wird deswegen gewählt.

2. Beispiel: Aufgabe des Treatments SF / Kulturfolger: „Verändere das Bild so, dass sich Tiere dort wohl fühlen.“ Der Ausdruck „wohl fühlen“ wird verwendet als strukturelle didaktische Reduktion (Killermann 1995, 255 ff.; Eschenhagen et al. 2001, S. 46 f.; Berck 2003) bzw. didaktische Transformation (vgl. Eschenhagen et al. 2001, S. 46 f.) für das eigentlich gemeinte „ökologisch intakt sein“. Für die Verwendung von „wohl fühlen“ sprechen die Verständlichkeit für Kinder, Prägnanz, Kürze und die gleichzeitige positive Attribuierung von „Natur“. Gegen diesen Ausdruck kann man die fehlende begriffliche Präzision, die aber das kindliche Konzept von „Ökologisch-Intakt-Sein“ übersteigen dürfte, und eine mögliche Förderung anthropomorph geprägter Vorstellung einer „natürlichen Idylle“ anführen. Dem zuletzt Genannten liegt der Gedanke

eines Animismus, einer „Beseelung nicht-menschlicher“ Objekte, zugrunde, der eine affektive Beziehung zur Welt offenbart. Gebhard (1990, S. 38 ff.) spricht sich nicht für einen Abbau dieser „Beseelung der Dinge“ aus. Darum wird der Ausdruck „wohl fühlen“ nicht vermieden.

2.3.4. Überlegungen zum „Erwartungshorizont“

Für alle Stationen lassen sich zu Treatments und impliziten / expliziten Lernzielen folgende Regelmäßigkeiten festhalten.

Für Treatment F kann die Erreichung des Grobziels_{ges} anhand der instruktionalen Vorgaben nicht unbedingt erwartet werden. Die Organisation des Unterrichts und der inhaltliche museale Rahmen erlauben aber eine (Mit-)Erreichung dieses Ziels. Es wird darum als implizites Lernziel bezeichnet. Das Grobziel_F steht im Fokus der Instruktion. Darum kann mit einiger Wahrscheinlichkeit seine Erreichung erwartet werden. Eine Operationalisierung dieses explizit angestrebten Lernziels ist sehr gut möglich.

Treatment SF erlaubt ähnlich wie beim Treatment F die Erreichung des Grobziels_{ges} anhand der instruktionalen Vorgaben nur bedingt. Darum gilt es als implizites Lernziel. Grobziels_{ges} kann jedoch erreicht werden und das mit größerer Wahrscheinlichkeit als bei Treatment F. Das Grobziel_{SF} steht trotz des offenen Antwortformates der hier verwendeten Aufgaben im Fokus der Instruktion und gilt darum als explizites Lernziel.

Treatment S: Das Grobziel_S entspricht dem Grobziel_{kog}, ist somit als explizites Lernziel zu bezeichnen. Die Offenheit der Aufgaben jedoch erlaubt kaum mehr Operationalisierung als implizite Lernziele beim Treatment F bzw. SF. Bei der im Treatment S für die Feinzielformulierung bezüglich der idealisierten Biotope (Urwald, Wacholderheide und Sandgrube) verwendeten Formulierung „Die Schüler sollen Inhalte (...) zum Ausdruck bringen können.“ kann man kaum noch von einem Feinziel sprechen. Die Operationalisierung bleibt nur formal erhalten. Inhaltlich steht dieses Lernziel eher auf der Stufe eines Grobziels. Für das Treatment S lässt sich kein präziseres Feinziel, dessen Erreichung mit einiger Wahrscheinlichkeit vorhergesagt werden

kann, formulieren. Die Aufgabenstellung ist so offen und lässt dem Schüler so viele Möglichkeiten, dass dies (eine präzise Formulierung operationalisierbarer Feinziele) scheitern muss. An diesem Punkt wird die Grenze der Kompatibilität traditionellen Vorgehens und des gemäßigt konstruktivistischen Ansatzes überschritten.

2.3.5. Formale und inhaltliche Charakterisierung der Treatmentaufgaben

2.3.5.1. Formale Charakterisierung

Mit Hilfe folgender begrifflicher Konzeptionen werden die Arbeitsaufträge für die Schüler charakterisiert (vgl. Anhang 2): Die Unterscheidung der Aufgaben erfolgt zunächst nach ihrer Zielsetzung, nämlich ob *eine* bestimmte Lösung richtig ist und diese während des Unterrichts im Museum gefordert und erwartet werden kann, hervorgebracht durch konvergentes Denken der Schüler, oder ob es eine Vielzahl gleichwertiger, vielleicht sogar gleichzeitig zu erwartender Lösungen gibt, hervorgebracht durch divergentes Denken der Schüler. Diese Unterscheidung zwischen einer richtigen Antwort (konvergent) und eines Spektrums gleichwertiger Antworten (divergent) wird in Anlehnung an König & Riedel (1971, S. 80 ff.) und ihrer Unterscheidung interner Operationen getroffen. Lienert (1969; S. 25 ff.) differenziert zwischen gebundenen und freien Aufgabenbeantwortungen. Bei den gebundenen Antworten gibt es festgelegte Möglichkeiten, für die sich der Bearbeiter zu entscheiden hat, bei den freien liegt die Art der Beantwortung in dessen Ermessen, also etwa verbal oder nichtverbal, z.B. zeichnerisch. Die Beantwortung kann mit den Denkopoperationen korrespondieren: Beispielsweise bedingen gebundene Beantwortungen vorherige konvergente Denkopoperationen. Eine weitere Kategorisierung folgt den hier gut verwendbaren Begrifflichkeiten nach Bortz & Döring (1995) zur Beschreibung von Testitems, auch wenn an dieser Stelle die Aufgaben des Treatments und nicht die Items des Fragebogens behandelt werden. Detaillierte Charakterisierung vergleiche Anhang 2).

2.3.5.2. Inhaltliche Charakterisierung

Aufgabenanalyse und methodische Implikation stehen in engem Zusammenhang mit dem jeweiligen Inhalt. Die nun folgende inhaltliche Charakterisierung wird anhand der Treatmentaufgaben bzw. der entsprechenden Lernziele durchgeführt.

Tabelle 1: Inhaltliche Analyse der Aufgaben der Treatments (Einzelheiten siehe Text).

	F		SF		S
Urwald	Schwarzspecht	<	Spechte	<	alle Inhalte
Grobziel _{kog}	implizit + Grobziel _F		implizit + Grobziel _{SF}		= Grobziel _S
Wa- cholderheid	Raubwürger	<	Würger	<	alle Inhalte
Grobziel _{kog}	implizit + Grobziel _F		implizit + Grobziel _{SF}		= Grobziel _S
Sandgrube	Eisvogel	≈	Eisvogel	<	alle Inhalte
Grobziel _{kog}	implizit + Grobziel _F		implizit + Grobziel _{SF}		= Grobziel _S
Archae- opteryx	Gemeinsamkeiten / Unterschiede von Archaeopteryx und Vogel bzw. Reptil	≈	Gemeinsamkeiten / Unterschiede von Archaeopteryx und Vogel bzw. Reptil	≈	Gemeinsamkeiten / Unterschiede von Archaeopteryx und Vogel bzw. Reptil
Grobziel _{kog}	= Grobziel _F		= Grobziel _{SF}		= Grobziel _S
Kulturland- schaft	Nützliches - von ei- nem Tier ausgehend (Feldhase)	≈ / ≠	Schädliches - von drei Tieren ausge- hend (Feldhase, Rebhuhn, Goldam- mer), Verbesserun- gen	≈ / ≠	Verbesserungen
Grobziel _{kog1} Grobziel _{kog2}	implizit = Grobziel _F		implizit = Grobziel _{SF}		implizit = Grobziel _S
Platzbedarf	Graphik, Platzbedarf, Vogelgewicht	≈	Platzbedarf, Vogel- gewicht, Graphik	≈	Platzbedarf von un- terschiedlichen Vö- geln
Grobziel _{kog}	= Grobziel _F		= Grobziel _{SF}		= Grobziel _S

Im Ideal sollten die in den Aufgaben der unterschiedlichen Treatments geforderten Inhalte konstant sein und sich nur in der Art des Unterrichts unterscheiden. Bei „Platzbedarf“ ist dies formal gegeben. Insbesondere bei den idealisierten Biotopen (Urwald, Wacholderheide, Sandgrube) kann diese Bedingung aufgrund der musealen Vorgaben gar nicht eingehalten werden, denn eine entsprechende Einschränkung würde dem Grundansatz der Studie widersprechen. In diesen Fällen wird als „zweitbeste“ Lösung im F-Treatment „ein Inhalt“, im SF-Treatment „eine Inhaltsgruppe“ und im S-Treatment „alle Inhalte“ in den jeweiligen Aufgaben zum Thema gemacht. Bei „Urwald im Fichtelgebirge“ bedeutet dies z.B. für das F-Treatment Aufgaben zum Schwarzspecht, zum SF-Treatment Aufgaben zu ausgewählten Beispielen zur Familie der Spechte (Picidae) und für das S-Treatment eine Aufgabe, die alle Inhalte des Dioramas umfasst. Für die Erhebung des Schülerwissens in den Fragebögen und vergleichbare Auswertung birgt dieses Vorgehen Schwierigkeiten (vgl. Material und Methoden: 4.2.1. Konstruktion des Fragebogens; Tabelle 1). Zur Ermittlung kognitiven Wissens werden darum unterschiedliche Messinstrumente eingesetzt.

Bei der Station „Platzbedarf“ wirken die methodischen Unterschiede so sehr in den wahrscheinlich von den Schülern v. a. wahrgenommenen Hauptinhalt hinein, dass dieser Einfluss thematisiert werden soll. Bei Treatment F werden die Aufgaben so gestellt, dass zwar der eigentliche Inhalt des Schaukastens (Platzbedarf von unterschiedlichen Vögeln) in jeder einzelnen Aufgabe gestreift wird, aber v. a. als Vehikel, um die Graphik zu lesen. Die kleinschrittigen Aufgaben lassen möglicherweise die graphische Darstellung zum eigentlichen Hauptinhalt werden. Aus diesem Grund wird hier die Graphik vor „Platzbedarf“ und „Vogelgewicht“ genannt. Der zu Grunde liegende Zusammenhang zwischen Körpergröße (bzw. –gewicht) und Platzbedarf könnte vom Schüler weniger wahrgenommen werden als die Art der Präsentation. Beim Treatment SF werden „Platzbedarf“ und „Vogelgewicht“ vor der Graphik genannt, da die Aufgaben viel eher einen Blick auf Inhalte und Zusammenhänge zulassen. Dennoch spielt der Umgang mit der graphischen Darstellungsform in jeder Teilaufgabe eine so entscheidende Rolle, dass auch hier die Graphik selbst als einer der Inhalte des Schaukastens genannt werden muss. Das Treatment S lenkt den Schüler auf das Verstehen und Selbsterklären des Inhalts des Schaukastens.

III. Entwicklung der Hypothesen

1. Hypothesen zur kognitiven Ebene

Die eingangs formulierte „zentrale Frage“ der Untersuchung, wodurch der größte Lernerfolg zu erreichen sei, durch Instruktion oder Konstruktion, muss nach Überlegungen zu theoretischen Grundlagen der Untersuchung in die Frage nach der richtigen „Balance“ zwischen Instruktion und Konstruktion für die Erreichung des größten Lernerfolgs korrigiert werden.

Nach der Entscheidung für ein bestimmtes Element, das für die Konstruktion (gewünschter) subjektiver Wirklichkeit besonders entscheidend ist, der Selbstbestimmung, ergeben sich methodische Beschränkungen und Möglichkeiten, die es erforderlich machen, die Eingangsfrage zu präzisieren: Welcher Grad konstruktivistisch orientierter bzw. instruktional orientierter „Bedingungen“, von Selbstbestimmung bzw. Fremdbestimmung, führt zu welcher Art von Lernerfolg? Allein auf kognitiver Ebene sind zwei grundlegend unterschiedliche Messinstrumente einzusetzen, die verschiedenartigen „Lernerfolg“ liefern, bzw. Einblicke aus mehr als einer Perspektive in das Resultat des Lernprozesses erlauben.

Daraus lassen sich die entsprechenden Hypothesen ableiten:

Ein hohes Maß an Fremdbestimmung führt zu „trägem Wissen“ (vgl. Gruber et al. 2000). Diese Form von Wissen ist durchaus tauglich, bestimmte Aufgaben zu lösen.

Hypothese eins (H₁): Die fremdbestimmt lernenden Schüler (F-Schüler) schneiden bei Items mit Antwortvorgaben (KogMit) besser ab als bei den Items ohne Antwortvorgeben (KogO).

Ein sehr hohes Maß an Selbstbestimmung führt bei Schülern, die dies nicht gewohnt sind, zu Überforderung und schlechten Leistungen. Die Schüler sind wahrscheinlich mit selbstbestimmtem Arbeiten wenig vertraut.

Hypothese zwei (H₂): Die kognitiven Ergebnisse der selbstbestimmt Lernenden (S-Schüler) fallen unterdurchschnittlich aus.

Nach Reinmann-Rothmeier & Mandl (2001) führt die „richtige Balance“ zwischen Instruktion und Konstruktion zum besten Lernerfolg. Dies gilt wahrscheinlich auch für die „richtige Balance“ zwischen Fremd- und Selbstbestimmung. Das SF-Treatment bietet diese Bedingung am ehesten.

Hypothese drei (H₃): Schüler der Mischform (SF-Schüler) schneiden im kognitiven Leistungstest am besten ab.

Die Bekanntheit eines außerschulischen Lernortes sollte zu besseren Vortestergebnissen führen. Damit starten diese „Museums-Kenner“ im Vergleich zu Nicht-Kennern auf einem höheren Wissensniveau.

Hypothese vier (H₄): Schüler, denen der Lindenhof vorher bekannt war, zeigen geringeren Lernzuwachs.

2. Hypothesen zu beiden Ebenen (kognitiv / affektiv)

Wie aus vielen empirischen Untersuchungen bekannt, z.B. Leibold (1997), Lude (2001), Randler (2004), muss man im kognitiven und affektiven Bereich mit geschlechtsspezifischen Unterschieden rechnen:

Hypothese fünf (H₅): Es gibt bei kognitiven und affektiven Lernleistungen geschlechtsspezifische Unterschiede.

3. Hypothesen zur affektiven Ebene

Die Analyse des Museums und die Charakterisierung des Unterrichts machen deutlich, wie entscheidend an unterschiedlichen Ansatzpunkten affektive Perspektiven der Rezeption sein können. Genannt seien Gries (1996) und Killermann (1995), die beide die Wichtigkeit affektiver Ebenen beim Besuch in einem Naturkundemuseum betonen, und Killermann (2000, 1995), Klautke & Köhler (1991) und Becker (2000), welche Notwendigkeit und Wert affektiver Dimensionen für die Umwelterziehung beschreiben.

Aus diesem Grund sollen affektive Ebenen bei der Evaluation des Lernerfolgs berücksichtigt werden.

„In der Biologiedidaktik fehlen Untersuchungen bezüglich der Lernemotionen (mit Ausnahme des kognitiv-emotionalen Konstrukts Interesse [vgl. Vogt 2000]) weitgehend“ (Randler 2004, S. 10). Randler referiert kurz die Arbeiten von Pondorf (1998a, zitiert nach Randler; vgl. Pondorf 1998b; beide zu Emotionen beim Computerlernen) und Hesse (2000, zitiert nach Randler; retrospektive Untersuchung an Erwachsenen zu früheren Erfahrungen im Biologieunterricht). Daneben könnte man Studien zu „Naturerfahrung und Naturschutzbewusstsein“, z.B. von Lude (2001, S. 61), der sich u. a. auf „sinnliche Wahrnehmung der Natur“ bezieht, anführen und die amerikanische Untersuchung von Thompson & Mintzes (2002), „Cognitive structure and the affective domain: knowing and feeling in biology.“ Diese Studie erfasst in „concept maps“ kognitive Konzepte zum Thema Haie und findet Korrelationen zwischen dieser kognitiven Ebene zur affektiven Einschätzung der Probanden, „attitudinal structure about sharks“. Thompson & Mintzes (2002, S. 646) nennen drei Aspekte („components“) von „attitude“: einen kognitiven Aspekt, der sich aus einer Reihe von Vorstellungen („series of beliefs“) zusammensetzt, einen affektiven, der v. a. die Gefühle, die ein Objekt oder Thema hervorruft („feelings stimulated by an (...) object“), umschreibt, und einen Aspekt, der Verhaltensprädispositionen bezüglich des „attitude objects“ beschreibt. Dieses Konzept lässt sich für die vorliegende Untersuchung in modifizierter Form nutzen: Gemessen werden sehr grundlegende Schülereinschätzungen zu Museum und Inhalten, wobei es nicht um eine differenzierte Beschreibung der „attitudes“, in dieser Untersuchung verstanden als „Einstellungen“ oder „Einschätzungen“, der Probanden zu bestimmten Themen geht, sondern zunächst um allgemeine Anhaltspunkte. Ähnlich wie Häußler et al. (1998, S. 150 und 159) die „affektive Leistung“ von Schülern oder „affektiver Wirkung“ einer unterrichtlichen Intervention beschreiben, soll in dieser Pilotstudie die Wirkung von Unterricht im Museum auf affektiver Ebene bestimmt werden. Dieses Konzept schließt die Akzeptanz der Probanden für die Intervention „Museumsbesuch“, als die „Chance, für bestimmte (...) Maßnahmen bei einer identifizierbaren Personengruppe (...) ausdrückliche oder stillschweigende Zustimmung zu finden“ (Lucke 1995, S. 104; vgl. Scharfenberg 2003, S. 29), ein. Zur affektiven Ebene sollen weitere Konzepte, die keinen klaren Bezug zu Ökologie oder Biologie haben, wie z.B. von Pekrun (1988), Hascher (1994), Hänze (1998) oder Büschges-Abel (1998), unberücksichtigt bleiben.

Ziel dieses Untersuchungsteils ist es, u. a. entscheidende Unterschiede der affektiven Rezeption der Probanden zwischen den Treatments quantitativ zu bestimmen. Während Reinmann-Rothmeier & Mandl (2001) diesem Faktor wenig Beachtung schenken, geht Cooper (1993, S. 17) bei konstruktivistischen Lernern von instrinsischer Motiviertheit aus. Damit sollte ein konstruktivistisch orientiertes Treatment zu positiverer affektiver Rezeption führen, als ein instruktionsbestimmtes Treatment. Die Einschätzungen zu Museumsbesuch und Stationen sollte positiver sein, die Akzeptanz höher. Darauf basierend wird folgende Hypothese aufgestellt:

Hypothese sechs (H₆): Die am deutlichsten konstruktivistisch orientierten Lerner (S-Schüler), bewerten den Museumsbesuch und seine Elemente besonders positiv.

IV. Material und Methoden

1. Gütekriterien der klassischen Testtheorie

Die vorliegende Untersuchung folgt der klassischen Testtheorie, die sich an naturwissenschaftlichen Messmodellen orientiert (Bortz & Döring 1995, S. 178 ff.). Damit wird für ein Testergebnis angenommen, dass es sich hierbei um den wahren Ausprägungsgrad eines Merkmals handelt, der von Messfehlern überlagert sein kann. „Die Kriterien der klassischen Testtheorie lassen sich (...) auf Fragebögen anwenden, die (...) die Funktion von Forschungsinstrumenten haben (...)“ (Bortz & Döring 1995, S. 179). Die Qualität solcher Forschungsinstrumente wird hauptsächlich durch drei Richtmaße gesichert: Objektivität, Reliabilität und Validität (Rost 1996, S. 31; Bortz & Döring 1995, S. 180; Lienert 1969, S. 12). Diese Gütekriterien sollen im Folgenden bezogen auf die durchgeführte Untersuchung dargelegt werden.

Zentrales Kriterium für die Objektivität einer wissenschaftlichen Untersuchung ist die intersubjektive Nachprüfbarkeit (verändert nach Popper 1984, S. 18 und Lienert 1969, S. 13; vgl. Zimbardo 1992, S. 14 f.). Nach Bortz und Döring (1995, S. 181 f.) sind Tests objektiv, „wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen, d.h. ein objektiver Test ist vom konkreten Testanwender unabhängig.“ Für Tests als Verfahren zur Untersuchung von Persönlichkeitsmerkmalen lässt sich diese Definition präzisieren: „Mit Objektivität ist gemeint, inwieweit das Testergebnis unabhängig ist von jeglichen Einflüssen außerhalb der getesteten Person (...)“, beispielsweise vom Versuchsleiter, der Stichprobenauswahl, der Itemkonstruktion, der Datenauswertung und allen denkbaren „situationalen Bedingungen“ (Rost 1996, S. 31). Damit deuten sich unterschiedliche Formen von Objektivität an. Bortz und Döring (1995, S. 180 f.) nennen Durchführungs-, Auswertungs- und Interpretationsobjektivität.

Durchführungsobjektivität ist gegeben, wenn das Testergebnis vom Untersuchungsleiter unabhängig ist (Bortz & Döring 1995, S. 180 f.). Um auf dieser Ebene den „Rosenthal-Effekt“, eine Beeinflussung der Untersuchungsergebnisse durch den Untersucher (Felser 2001, S. 457), zu vermeiden, wird die eigentliche Ausführung des Unterrichts im Museum nicht vom Untersuchungsleiter geleistet, sondern von elf Be-

treuerinnen, denen die expliziten Untersuchungshypothesen unbekannt sind. Da nicht nur den Probanden, sondern auch den Betreuerinnen die Untersuchungshypothesen nicht bekannt sind, könnte man nach Felser (2001, S. 457) von „Doppel-Blind-Versuchen“ sprechen. Die Betreuerinnen haben den größten Anteil an der praktischen Durchführung des Museumsbesuchs. Um ein möglichst hohes Maß an Genauigkeit in der Treatmentimplementierung zu erreichen, werden die Betreuerinnen vor dem Museumsbesuch umfassend und ausführlich geschult. Weiter gibt es Checklisten zur Durchführung (Anhang 3). Der Untersuchungsleiter ist an jedem Tag der Durchführung der Studie anwesend. Damit ist bezüglich der Treatments ein hohes Maß an Standardisierung gewährleistet und Durchführungsobjektivität gegeben.

Eine Beeinflussung der Untersuchung, die in der Wirkung der Persönlichkeit der Betreuerinnen auf die Schüler liegt, kann als Störvariable nicht ganz vermieden werden. Durch die relativ große Anzahl von elf unterschiedlichen Betreuerinnen wird das Gewicht dieses Einflusses verringert. Bezüglich eines Treatmentvergleichs spielt diese Störvariable keine Rolle, da auf eine Rotation der Betreuerinnen geachtet wird, so dass jede Betreuerin gleich oft jedes der drei Treatments begleitet (Anhang 4): Rotationsplan). (vgl. Kapitel Material und Methoden 6.)

Die Auswertung eines Tests soll von der Person des Auswerter unabhängig sein. Ist das gegeben, spricht man von Auswertungsobjektivität (Bortz & Döring 1995, S. 180). Wie weiter unten dargelegt ist (vgl. Material und Methoden 4.3.2.), kann man nach Diehl & Staufenbiel (2002, S. 161 f.) und Bortz & Döring (1995, S. 254) in dieser Untersuchung in allen sinnvoll zu testenden Fällen von guter Urteilerübereinstimmung und damit von akzeptabler Auswertungsobjektivität ausgehen.

Unter „Interpretationsobjektivität“ wird hier die „Einordnung eines Ergebnisses in den Kontext bekannter Befunde“ verstanden. Interpretationsobjektivität kann bei dieser Pilotstudie, in der keine standardisierten Tests Anwendung finden, nur im Rahmen vorhandener Sekundärliteratur, z.B. Grolnick & Ryan (1987), Mandl et al. (1992), Lord (1997), Hay (2001) und Randler & Bogner (2002), nicht aber wie Bortz & Döring (1995, S. 181) vorschlagen, mit Hilfe tabellarischer Vergleichsmaßstäbe aus Testhandbüchern gewährleistet werden.

Die Reliabilität gibt die Messgenauigkeit eines Tests an, definiert als Zuverlässigkeit bei wiederholter Messung unter identischen Bedingungen das gleiche Testergebnis zu erzielen (Rost 1996, S. 31). Gäbe es keine Messfehler, so wäre vollkommene Reliabilität gegeben (Bortz & Döring 1995, S. 181). Entscheidend ist eine Quantifizierung der Reliabilität, um das Maß der Reproduzierbarkeit von Messergebnissen einschätzen zu können. Eine Reliabilitätsanalyse incl. der Bestimmung von Trennschärfe und Schwierigkeitsgrad folgt unter Material und Methoden 4.3.1.

Die Validität („Gültigkeit“; Bortz & Döring 1995, S.185) eines Tests sagt aus, inwieweit er misst, was er messen soll, liefert damit Information über die Aussagefähigkeit eines Tests bezüglich seiner Intention (Rost 1996, S. 31). Eine Vermischung der Ebene der testtheoretischen Betrachtung von Validität, die die Qualität der Messinstrumente angibt, mit der Ebene der Deutung als Gütekriterium empirischer Untersuchungsdesigns, nämlich als interne und externe Validität (vgl. Material und Methoden 3. Stichprobe), sollte vermieden werden (Bortz & Döring 1995, S. 185). Es lassen sich bezüglich der testtheoretischen Betrachtungsebene der Validität drei Hauptarten unterscheiden: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität.

Inhaltsvalidität (= Augenscheinvalidität) gilt als gewährleistet, wenn das Item inhaltlich die wesentlichen Aspekte des zu messenden Konstrukts erfasst (Bortz & Döring 1995, S. 185). Dies wird in allen affektiven und kognitiven Teilen des Fragebogens nach Augenschein erfüllt. Ein Verfahren zur numerischen Bestimmung, z.B. in welchem Maße Inhaltsvalidität erfüllt ist, gibt es nicht (Bortz & Döring 1995, S. 185).

Kriteriumsvalidität gilt als gegeben, wenn ein zu testendes Merkmal mit einem korrespondierenden manifesten Merkmal (Außenkriterium) in hinreichendem Maße übereinstimmt. Die Resultate der Schüler in den kognitiven Teilen des Fragebogens sollten mit der kognitiven Leistung der Schüler in anderen Bereichen korrespondieren, z.B. der Schulleistung. Dies wurde überprüft (vgl. Material und Methoden 4.2.1 Konstruktion des Fragebogens). Die affektiven Testteile zeigen ein über die Schwierigkeit, ein reliables und valides Merkmal zu finden, hinausgehendes Problem auf. Es lässt sich kein angemessenes Außenkriterium finden, das hinreichend überzeugend mit den Messwerten korrespondiert. Innerhalb des Fragebogens jedoch werden mehrere ähnliche Merkmale erhoben, die sich aufeinander beziehen lassen (vgl. Material und Methoden 4.2.1 Konstruktion des Fragebogens).

Inwieweit Konstruktvalidität, eine in sich konsistente und widerspruchsfreie Einbindung in die Theorie, gestützt durch schon bekannte Forschungsergebnisse, gegeben ist, kann erst in der Diskussion geklärt werden, wobei schon jetzt auf ein Dilemma hingewiesen sei. Entsprechen Ergebnisse nicht den vorherrschenden Theorien bzw. den aus ihnen abgeleiteten Hypothesen, so ist die Bedeutung des Ergebnisses nicht festzulegen. Es kann entweder bedeuten, dass die Hypothese zu verwerfen ist oder dass keine Konstruktvalidität gewährleistet ist.

2. Vorstudien

Vor der hier vorgestellten Hauptuntersuchung gab es eine Reihe von Vorstudien. Als erste explorative empirische Arbeiten sind eine Seminararbeit und zwei Examensarbeiten zum Unterrichten in den beiden Museen des Umweltschutz-Informationszentrums Lindenhof zu nennen, die halfen, inhaltliche Fragen aufzuwerfen und erste methodische Ansätze anzubahnen.

Im Folgenden gab es drei Voruntersuchungen, die zielgerichtet auf die Hauptuntersuchung hinführten. In die erste und zentrale Voruntersuchung gehen die Daten von vier Schulklassen und 89 Schülern ein. Auf die Darstellung dieser Ergebnisse soll in Anbetracht des Umfangs der Hauptuntersuchung verzichtet werden. Es sei verwiesen auf Wilde & Klautke (2002) und Wilde et al. (2003). Der Fragebogen für die Hauptuntersuchung wurde daraufhin justiert, z.B. wurde sein Hauptteil (Items mit Antwortvorgabe / kognitiv) von 27 auf 18 Items reduziert. Aufgegeben wurden Items, die nicht in die definierte Range des Schwierigkeitsgrades von 10 bis 90% nach Leibold (1997) fielen, und die Items der geringsten Trennschärfe. Des Weiteren wurden Treatments und Zeitvorgaben einander angeglichen. Diese korrigierten Treatments und das verbesserte Messinstrument wurden an einer Schulklasse einer rein organisatorisch-methodisch ausgerichteten Erprobung unterzogen (zweite Voruntersuchung).

Eine letzte Voruntersuchung mit 53 Schülern aus zwei Klassen galt der Einschätzung des Pretest-Effekts (Bortz & Döring 1995, S. 472), exemplarisch durchgeführt bezüglich der Items mit Antwortvorgabe des kognitiven Teils des Fragebogens in der endgültigen Fassung (*KogMit18*, vgl. Material und Methoden 4.1.).

Zuvor wurde die Wahrscheinlichkeit, mit der ein bayerischer Gymnasiast der fünften Jahrgangsstufe aufgrund seines Vorwissens diese Anforderungen meistern kann, eingeschätzt. Dies erfolgte mit Hilfe der Erfahrung von Experten, für diese Aufgabe naturgemäß bayerische Gymnasiallehrer. Von den beiden Lehrern waren keine „ihrer“ Klassen an der Untersuchung beteiligt, so dass durch Sympathie, Antipathie, mittelbare Selbstdarstellung o.ä. keine Verfälschungen zu erwarten sind. Entsprechend hoch fällt der Wert für die Signierobjektivität ($\kappa=0,769$) aus, der eine hohe Urteilerkonkordanz anzeigt (Diehl & Staufenbiel 2002, S. 161 f.) und damit akzeptable Objektivität ausdrückt (Bortz & Döring 1995, S. 254). Das Ergebnis der Lehrer ist eindeutig: Schüler der fünften Klasse können diese Items aufgrund ihres Vorwissens wahrscheinlich zum größten Teil nicht beantworten (Einzelheiten siehe Anhang 5).

Zur Überprüfung des Pretest-Effektes: Das Testdesign folgt dem Schema „Vortest – kein Treatment – Nachtest I – Nachtest II“ (anstatt: „Vortest – Treatment – Nachtest I – Nachtest II“). Wie zu erwarten, zeigen sich zwischen Vortest, Nachtest I und II im Mittel nur sehr geringe Unterschiede. Die Differenzen betragen weniger als 0,5 richtige Antworten bei 18 Items und sind nicht statistisch bedeutsam (Friedmann-Test, $p=0,226$). Die exakten Ergebnisse finden sich im Anhang 6.

Anscheinend ist der Pretest-Effekt für diese Untersuchung von geringer Bedeutung. Auf eine Kontrolle durch den „Solomon-Vier-Gruppen-Plan“ nach Bortz & Döring (1995, 503 f.) kann darum verzichtet werden. Stattdessen werden in der Hauptuntersuchung drei Items mit museumsfremden Inhalten in den Fragebogen aufgenommen. Inhaltlich behandeln die drei Items neuseeländische Vögel. Ziel ist es, zu überprüfen, ob es in der Hauptuntersuchung so etwas wie einen „Carry-Over-Effekt“ im Sinne einer Verbesserung der Testleistung allein durch wiederholtes Bearbeiten des Testes gibt. Die Ergebnisse bezüglich dieser „Carry-Over“-Items unterscheiden sich zwischen Vortest und Nachtest I nicht (Wilcoxon-Test, $p=0,946$). Damit kann man für die Hauptuntersuchung einen Einfluss durch Pretest-Effekte vernachlässigen.

3. Stichprobe

Zuerst stellt sich die Frage nach dem Umfang der Stichprobe. In der Voruntersuchung zeigt sich, dass selbst der Museumsbesuch als solcher für den Lernerfolg der Schüler mit einer Effektstärke von $\omega^2=0,50$ „nur“ etwa die Hälfte des Lernerfolgs erklärt (vgl. Wilde et al. 2003, S. 130). Die Effektstärke der Treatmentunterschiede bezüglich des Lernerfolgs liegt weit niedriger. Dies kann kaum überraschen, denn ein so komplexes Geschehen wie Unterricht lässt sich nicht monokausal erklären. Eine präzise und verlässliche Einschätzung der Stichprobengröße bei diesen zu erwartenden kleinen bis mittleren Effekten und bei nicht einzuschätzenden weiteren Einflüssen, die zur Ausprägung relevanter Merkmale erheblich beitragen können und so eine weitere Gruppierung erforderlich machen können, ist auch mit den Hilfen zur Einschätzung der optimalen Stichprobenumfänge nach Bortz & Döring (1995, S. 575 ff.) kaum möglich. Aus der oben zitierten Voruntersuchung ist bekannt, dass sich trotz z. T. geringer Effektstärken statistisch bedeutsame Ergebnisse ermitteln lassen. Gleichzeitig ist sehr wahrscheinlich, dass weitere Differenzierungen der Gruppen weitere Ergebnisse zu Tage fördern könnten. Dieser Konstellation geschuldet, wird eine möglichst große Stichprobe gewählt, die aber für die Treatmentdurchführung und die Auswertung noch gerade beherrschbar ist.

Inwieweit diese Stichprobe repräsentativ sein kann, wirft grundsätzliche Probleme auf. Repräsentativität ist gegeben bei möglichst großer Ähnlichkeit, idealerweise Gleichheit, der Merkmale der Probanden einer Stichprobe mit den Merkmalen einer Population (vgl. Bortz & Döring 1995, S. 371). Das ist zu erreichen durch Zufallsauswahl einer genügend großen Anzahl von Probanden. Aus praktischen Gründen ist eine echte Zufallsauswahl nicht realisierbar.

In der Hauptuntersuchung sind bei 366 von 385 Schülerinnen und Schülern aus 14 Schulklassen die Fragebögen für die Auswertung nutzbar, da diese Probanden am Treatment und an allen drei Tests teilnahmen. Das ist eine Quote von über 95%. Die 14 Schulklassen aus der fünften Jahrgangsstufe rekrutierten sich aus vier Gymnasien, zwei aus Bayreuth und zwei aus dem Umland. Das Alter der Schüler lag zum Zeitpunkt der Untersuchung im Durchschnitt bei 10,8 Jahren. Tabelle 7 liefert eine bilanzierende Darstellung weiterer soziodemographischer Daten. Zuerst zur letzten

Kategorisierung, den Probanden, denen der Lindenhof vor dem organisierten Schulbesuch nicht bekannt war (Lindenhof-Nicht-Kenner), und den Probanden, die das Umweltschutz-Informationszentrum vorher schon besucht hatten (Lindenhof-Kenner). Später zeigt sich, dass diese Größe von Bedeutung ist, so seien schon hier diese beiden Typen unterschieden. Die Mehrzahl der Probanden kennt den Lindenhof vor dem Museumsbesuch noch nicht.

Tabelle 2: Soziodemographische Daten. Erläuterungen im Text.

	Σ auswertbarer Probanden	Unzureichende Angaben	Teil der Untersuchung versäumt
Gesamtgruppe	366	-	19
Jungen / Mädchen	188 / 177	1	
Lindenhof-Nicht-Kenner / Lindenhof-Kenner	241 / 121	4	

Eine präzisere Charakterisierung der Schulen und der Schüler, die eine genauere Einordnung des gesellschaftlichen Umfelds erlauben würde, kann nicht erfolgen, da die Verpflichtung zur anonymen Behandlung der Daten – auch auf der Ebene einzelner Schulen – zugesichert werden musste. Die Untersuchung fand während der Unterrichtszeit statt und nahm incl. Transportzeiten, Pausen, etc. einen gesamten Schultag in Anspruch.

Die Auswahl der Stichprobe entscheidet über die Klassifizierung als experimentelle oder quasi-experimentelle Untersuchung und damit über Validität und Gültigkeitsanspruch der Untersuchungsbefunde. Hier sind zwei Ebenen zu unterscheiden.

Die erste Ebene betrifft die Auswahl der Schulen und Schüler für die Untersuchung. Es wurden nur Gymnasien aus Bayreuth und dem näheren Umland berücksichtigt. Und auch innerhalb dieser beschränkten Auswahl war es aus organisatorischen Gründen nicht möglich, alle verbleibenden Schüler der fünften Jahrgangsstufe einem Randomisierungsverfahren zu unterziehen und die ermittelte Auswahl an Schülern in der benötigten Stichprobenmenge in das Naturkundemuseum einzuladen. Trotz einer Genehmigung des Kultusministeriums (Anhang 7) konnten zum einen nur Gymnasien einbezogen werden, deren Direktoren die Teilnahme an der Untersuchung be-

willigten, zum anderen kamen pro Museumsbesuch aus schulorganisatorischen Gründen jeweils alle Schüler einer Klasse. Diese der realen Lebenswirklichkeit geschuldeten Einschränkungen erhöhen eher die Übertragbarkeit der Untersuchungsergebnisse auf andere ebenfalls im „wirklichen Leben“ stattfindende Museumsbesuche und so die externe Validität. Damit handelt es sich jedoch bei der vorliegenden Auswahl an Probanden um natürliche Gruppen. Das Kriterium einer Randomisierung ist nicht erfüllt. Damit kann es sich qua definitionem nur um eine quasi-experimentelle Untersuchung handeln. Das verringert die interne Validität, d.h. die Anzahl plausibler Alternativerklärungen bezüglich der zu untersuchenden Hypothese steigt (Bortz & Döring 1995, S. 52 ff.).

Die zweite Ebene betrifft die Zuordnung der Probanden zu den Treatmentgruppen. Auf dieser Ebene wurden personenbezogene Störvariablen durch eine zufällige Aufteilung in die Treatmentgruppen ausgeschaltet. Diese Randomisierung erhöht die interne Validität. Einer Generalisierung des Untersuchungsergebnisses widerspricht es jedoch nicht, so dass die externe Validität dadurch nicht verringert wird. Auf dieser Ebene würde man nicht von einer quasi-experimentellen sondern eher von einer experimentellen Untersuchung sprechen.

Entscheidend ist an dieser Stelle nicht die semantische Klärung der Klassifizierung der Untersuchung sondern die Erreichung möglichst hoher interner und externer Validität. Hierzu sei ein Exkurs erlaubt, der über die Bedeutsamkeit dieses Punktes für die Stichprobe hinausgeht. Interne Validität lässt sich durch die Eliminierung von Alternativerklärungen erhöhen, zu erreichen durch die Vermeidung der Beeinflussung von Testergebnissen durch jegliche störende Variable, seien es personenbezogene, z.B. besonders intelligente Probanden, seien es äußere Einwirkungen, z.B. Straßenlärm bei der Bearbeitung eines Tests. Kurz: Die interne Validität steigt mit einer Zunahme an Standardisierung. Das Museum bietet den günstigen Fall ähnlich einer Laborsituation (fast) alle beeinflussenden Faktoren – zumindest die nicht-personenbezogenen – kontrollieren zu können. Die externe Validität erhöht sich, wenn die gefundenen Ergebnisse über die Untersuchungssituation und die teilnehmenden Probanden hinaus zu verallgemeinern sind. Das ist durch möglichst „realitätsnahe“ Untersuchungssituationen zu erreichen. Genau das, nämlich authentische Situationen, bietet ein Museum. Es stellt kein Lernlabor dar, sondern einen den

Schülern grundsätzlich bekannten und auch außerhalb der Untersuchung genutzten außerschulischen Lernort. Oft beeinflussen Korrekturen zugunsten der internen Validität die externe Validität negativ (und umgekehrt) (Bortz & Döring 1995, S. 52). Dies trifft bei dem Untersuchungsort Museum nur zum Teil zu, denn einerseits bietet es viele Möglichkeiten der Standardisierung, andererseits Authentizität. Wie oben bereits ausgeführt, wurde versucht, interner wie externer Validität im Rahmen der organisatorischen Gegebenheiten so weit wie möglich Rechnung zu tragen.

4. Erhebungsinstrument / Fragebogen

Als wichtigstes methodisches Instrument dieser Untersuchung wird ein Fragebogen verwendet, der nach Bortz & Döring (1995, S. 175) als ein „Psychologischer Test“ aufgefasst werden kann. Lienert (1969, S. 7) versteht unter dem Begriff Test „ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.“ Die Ergebnisse der Untersuchung werden demnach Persönlichkeitsmerkmale der Probanden sein, die durch mathematisch-statistische Verfahren abzusichern sind und damit Aussagen zu den drei Treatments und weiteren Variablen zulassen werden.

4.1. Formale Beschreibung

Für die hier gestellten Untersuchungsfragen bezogen auf die Museen des Umweltschutz-Informationszentrums Lindenhof gibt es keine standardisierten Tests. Darum müssen informelle Tests entwickelt werden. Verwendet wird ein Messwiederholungsverfahren: Zu drei Messzeitpunkten wird jeweils ein Fragebogen eingesetzt, einer 7 ± 2 Tage vor (Vortest) und einer direkt nach dem Treatment (Nachtest I), ein dritter nach etwa 40 ± 2 Tagen (Nachtest II, vgl. Tabelle 3). Die Fragebögen sind mit dem Geburtsdatum der Schülerin / des Schülers und der zwei ersten Buchstaben des Vornamens ihrer / seiner Mutter gekennzeichnet. Damit sind sie anonymisiert, und es lassen sich die drei Bögen vom selben Probanden einander eindeutig zuordnen. Soziodemographische Daten werden nur im Vortest erhoben, affektive Einschätzungen zum Museumsbesuch nur in den beiden Folgetests, Nachtest I und II. Die übrigen Teile des Fragebogens sind inhaltlich und formal identisch (vgl. Tabelle 3).

Im Fragebogen werden den kognitiven und den affektiven Bereich betreffend „Items mit und ohne Antwortvorgabe“ verwendet. Dieser Terminus beschreibt sehr einfach und klar die verwendeten Itemtypen und wird im weiteren Text beibehalten. Dennoch sollen an dieser Stelle zum Zweck der exakten sprachlichen Verortung im Sinne der empirischen Sozialforschung alle Items den in Bortz & Döring (1995) und den in Lienert (1969) verwendeten Begrifflichkeiten zugeordnet werden.

Zuerst zum kognitiven Bereich (vgl. Tabelle 3). Im zentralen Teil des Fragebogens *KogMit25* werden nach Lienert (1969, S. 25) „Items mit gebundener Aufgabenbeantwortung“, nach Bortz & Döring (1995, S. 194 f.) „Multiple-Choice-Aufgaben“ bzw. „Items mit Antwortvorgaben“ oder – für diesen Itemtyp noch präziser – Items mit „Alternativantworten“ verwendet. Der Abschnitt *KogO* bedient sich nach Lienert (1969, S. 28 f.) „Items mit freier Aufgabenbeantwortung“, genauer von „Kurzaufsatzaufgaben“, nach Bortz & Döring (1995, S. 194) „Items mit offener Beantwortung“. Aus formalen Gründen ist die Erhebung der soziodemographischen Daten im Vortest ebenfalls hier anzuführen. Die Typen variieren sehr stark von Items mit „Alternativantworten“ (Bortz & Döring 1995, S. 195) / „Items mit gebundener Aufgabenbeantwortung“ (Lienert 1969, S. 25) über Items mit „halboffener Beantwortung“ (Bortz & Döring 1995, S. 194 f.) / „Ergänzungsaufgabe“ (Lienert 1969, S. 28) und Items mit „offener Beantwortung“ (Bortz & Döring 1995, S. 194) / „Kurzaufsatzaufgaben“ (Lienert 1969, S. 28 f.).

Die wesentlichen Bereiche der Tests zur affektiven Ebene *AffMu*, *AffMuUr* und *AffMuPla*, außerdem „*Tieraffektivität*“ bestehen aus „Multiple-Choice-Aufgaben“ (Bortz & Döring 1995, S. 194 f.) / „Items mit gebundener Aufgabenbeantwortung“ (Lienert 1969, S. 25) mit fünfstufigen Ratingskalen (vgl. Likert nach Bortz & Döring 1995, S. 203 f. und S. 164), so wie sie von Probanden am häufigsten präferiert werden (Bortz & Döring 1995, S. 167), jeweils erweitert um die Möglichkeit einer Nicht-Entscheidung. Im Teil *Freie Äußerung* gibt es ein Item mit „Alternativantworten“ (Bortz & Döring 1995, S. 195) / „Item mit gebundener Aufgabenbeantwortung“ (Lienert 1969, S. 25) und zwei Items mit „offener Beantwortung“ nach Bortz & Döring (1995, S. 194), nach Lienert (1969, S. 28 f.) „Items mit freier Aufgabenbeantwortung“; eines dieser Items ist eine „Kurzaufsatzaufgabe“.

Tabelle 3: Übersicht über die Fragebögen. Die kognitiven Teile (KogMit, KogO) sind durch Doppelrahmen hervorgehoben. Die im weiteren Text verwendeten Bezeichnungen für die jeweiligen Testteile sind unterstrichen. Die Items werden exemplarisch angedeutet (*kursiv*). Im Anhang 8 bis 15 findet sich die vollständige Darstellung dieses Messinstruments; weitere Erläuterungen im Text.

Fragebogen	Vortest	Nachtest I / II	Bezeichnung
1. Seite	Identifizierung, soziodemographische Daten (5Items)		Soziodemographisch (Vortest)
1. Seite		Identifizierung + affektiv, Items mit Antwortvorgaben (8 Items) <i>„So gut hat mir ... gefallen!“ (Likert)</i>	AffMu (<i>affektiv / Museum, allgemein</i>) (Nachtest I und II)
2., 3. Seite	„Indoor-Outdoor“ (1 Item), sonst wie rechts	affektiv, Items mit Antwortvorgaben (8 Items) <i>„Empfindest Du gegenüber den folgenden Tieren Zuneigung oder Abneigung?“ („quasi“-Likert)</i>	<u>„Tieraffektivität“</u> (Vortest, Nachtest I und II)
4. Seite	wie rechts	kognitiv, Items ohne Antwortvorgabe (6 Items) <i>„Wie sieht für Dich der Urwald des Fichtelgebirges aus?“ (offene Antwort)</i>	Nebentest: <u>KogO</u> (<u>kognitiv, Items ohne Antwortvorgaben</u>)
5. Seite	-	affektiv, Items mit Antwortvorgaben (7 Items) <i>„So gut hat mir ... gefallen!“ (Likert)</i>	AffMuUr (<i>affektiv / Museum (Urwald)</i>) (Nachtest I und II)
6. Seite	-	affektiv, Items mit Antwortvorgaben (4 Items) <i>„So gut hat mir ... gefallen!“ (Likert)</i>	AffMuPla (<i>affektiv / Museum (Platzbedarf)</i>) (Nachtest I und II)
7., 8., 9. Seite	wie rechts	kognitiv, Items mit Antwortvorgabe (18 + 7 Items) <i>„Ein Grünfink braucht mehr Platz als ein Kernbeißer!“ (Antwortmöglichkeiten: stimmt, stimmt nicht, weiß nicht“)</i>	Haupttest: <u>KogMit18 bzw. 25</u> (<u>kognitiv, Items mit Antwortvorgaben</u>)
10. Seite	-	affektiv, Items mit / ohne Antwortvorgabe (3 Items) <i>„Willst Du zu Deinem Besuch noch etwas sagen?“ (offene Antwort)</i>	Freie Äußerung (verbal, zeichnerisch) (Nachtest I und II)

4.2. Inhaltliche Charakterisierung

Der verwendete Fragebogen berührt zwei inhaltliche Dimensionen, nämlich die eines Leistungs- und die eines Persönlichkeitstests (Bortz & Döring, 1995, S. 175 ff.). Ein Leistungstest ist der Fragebogen immer dann, wenn es um die Beantwortung von Aufgaben objektivierbaren Inhalts geht, also bei den Items mit und ohne Antwortvorgabe der kognitiven Ebene (*KogO*, *KogMit*). Ein Persönlichkeitstest ist der Fragebogen, wenn objektive Beurteilungen keine Rolle spielen, also bei den Items mit und ohne Antwortvorgabe der affektiven Ebene (*AffMu*, *AffMuUr*, *AffMuPla*, „*Tieraffektivität*“ und *Freie Äußerung*). Bisweilen sind Überlappungen dieser Dimensionen zu erwarten, wenn beispielsweise die Probanden bei offenen Fragen zur kognitiven Ebene (*KogO*), Items der ersten Dimension, affektiv gefärbte Aussagen machen, die dann auch unter dem Aspekt dieser zweiten inhaltlichen Dimension, nämlich der eines Persönlichkeitstests, ausgewertet werden können.

4.2.1. Konstruktion des Fragebogens

Affektive Einschätzungen zum Museumsbesuch lassen sich nur im Nachtest I und II erheben. Vorher ist das Museum bei den Probanden i. d. R. noch nicht bekannt. Die Schüler bewerten mit den Items zu einer allgemeinen Einschätzung des Museumsbesuchs und der Stationen (incl. Kindermuseum) *AffMu* eine Dimension. Dies belegt eine Faktorenanalyse (Hauptkomponenten-Analyse, SPSS Version 11.5). Nur ein Faktor wird identifiziert (vgl. Anhang 16). Die Überprüfung der Faktorstabilität ergibt 0,92 (vgl. Bortz 1999).

Für die Prüfung der Kriteriumsvalidität lässt sich bezüglich der affektiven Schülereinschätzungen kein überzeugendes Außenkriterium finden. Jedoch positive Korrelationen innerhalb von *AffMu*, nämlich des ersten Items (Bewertung des gesamten Besuchs) gegen die übrigen Items (Bewertung der Stationen), sprechen für die Validität der Items (vgl. Anhang 17). Etwas globaler betrachtet, kann man *AffMuUr*, *AffMuPla* als weitere Items bzw. Testteile mit ähnlicher Zielrichtung sehen, wodurch die positiven Korrelationen zu *AffMu* als weitere Bestätigungen der Kriteriumsvalidität aufzufassen sind (vgl. Anhang 18). Bezogen auf eine affektive Mischbewertung von Museumsbesuch und –inhalten misst der Test durchaus, was er zu messen vorgibt.

In einem weiteren Item werden die Schüler befragt über eine in die Zukunft weisende Intention als Konsequenz des gerade Erlebten, nämlich die Absicht, das Museum ein weiteres Mal zu besuchen (*Freie Äußerung*, erstes Item). Daneben wird den Probanden die Möglichkeit einer völlig freien Bemerkung gegeben, die sie häufig für eine Gesamtbewertung des Museumsbesuchs nutzen (*Freie Äußerung*, zweites Item).

Gleichwohl sind Items zur affektiven Ebene, die sich – im Gegensatz zu den obigen Fragebogenteilen - auf Nachtests und Vortest beziehen lassen, und somit Vortest-Nachtest-Vergleiche erlauben, interessant. Dadurch wäre der Einfluss des Treatments auf die affektive Bewertung der Probanden zu ermitteln. Dies wird versucht, indem die Probanden vor und nach dem Museumsbesuch zu „Objekten“ befragt werden, die ihnen schon vor dem Museumsbesuch bekannt sind. Es handelt sich um „Objekte“, die in den Treatments im Museum eine Rolle spielen und Objekte, die nicht im Museum zu sehen sind (Blindprobe). Bei entsprechenden Veränderungen der affektiven Einschätzungen der Probanden zu „Museums-Objekten“ nach den Treatments (und fehlender Veränderung bei „Nicht-Museums-Objekten“), wären für diese Effekte die Treatments ursächlich.

Folgendes wird im Fragebogenteil „*Tieraffektivität*“ versucht, wobei vorzuschicken ist, dass dieser Testteil als orientierende Vorstudie zu sehen ist. Weder Pretest-Effekte können ausgeschlossen werden, noch haben Itemanalysen und entsprechende Optimierungen stattgefunden. In dieser Teilstudie wird die affektive Einschätzung zu 14 Tieren, die Schülern der fünften Jahrgangsstufe bekannt sein dürften, in Vor- und Nachtests in 14 Items erfragt. Veränderungen der Schülerbewertungen sollten auf die Treatments zurückzuführen sein. Weitere Erläuterungen und die Daten zu dieser Erhebung finden sich im Anhang 19.

Von ganz anderer Art ist das letzte Item des Teils *Freie Aufgaben* und damit das letzte Item des gesamten Fragebogens: „(...) zeichne (...), was Dir am Lindenhof am allerbesten gefallen hat.“ Die einfache Instruktion genügt dieser sehr robusten Erhebungstechnik in allen Altersklassen (White & Gunstone 1992, S. 98 ff.). Zeichnen eignet sich besonders für die Untersuchung affektiver Ansichten. „The drawers feelings about the subject may appear, (...)“ (White & Gunstone 1992, S. 101).

In den die affektive Ebene erfassenden Testteilen *AffMu*, *AffMuUr*, *AffMuPla* und „*Tieraffektivität*“ werden fünfstufige Likert-Skalen verwendet. Diese ungeradzahliges Rating-Skalen erzwingen keine Entscheidungen zu Gunsten einer der Polarisierungen, da hier eine neutrale Einschätzung möglich ist. Fünfstufige Skalen werden von Urteilern am häufigsten präferiert (Bortz & Döring 1995, S. 167). Sie erlauben abgestufte Entscheidungen, überfordern die Differenzierungsfähigkeit der Probanden jedoch nicht. Die Güteeigenschaften Reliabilität und Validität sind von der Wahl der Skala, geradzahlig oder ungeradzahlig, unabhängig (Bortz & Döring 1995, S.167).

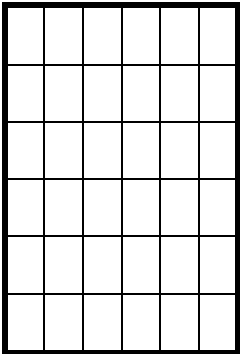
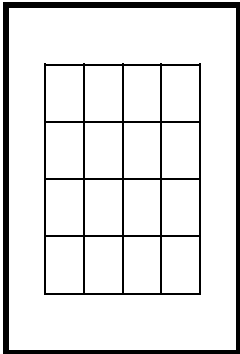
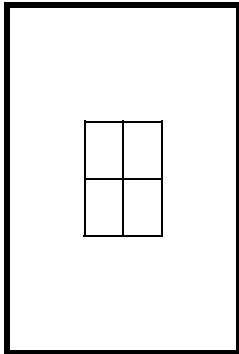
Die kognitiven Teile des Fragebogens, *KogO* und *KogMit*, haben zum Ziel, das Ausgangsniveau des Wissens der Schüler *vor* und deren kognitive Veränderungen *nach* dem Museumsbesuch zu beschreiben. Sie wurden entwickelt im Wissen um folgendes Dilemma, das dem zu Grunde liegenden theoretischen Konstrukt immanent ist. In den Treatments wird in erster Linie die Selbststeuerung der Schüler variiert. Daraus resultiert unterschiedliche Vorhersagbarkeit des eigentlich im Museum bearbeiteten Inhalts bezüglich der drei Treatments. Das kann die kognitiven Ergebnisse beeinflussen, mit Hilfe derer jedoch die Selbststeuerung beurteilt werden soll. Diese „Rückkopplung“ ist nicht völlig zu beheben.

Die Unterschiedlichkeit der Treatments äußert sich bei den F-Probanden u. a. durch relativ leichte Vorhersagbarkeit des wahrscheinlich im Museum bearbeiteten Inhalts, da die Schüler klare instruktionale Vorgaben haben. Bei den Probanden des Treatments SF ist die Vorhersage der tatsächlich bearbeiteten Inhalte schlechter möglich, bei den Probanden des Treatments S dagegen ist eine Vorhersage sehr problematisch, da diese Schüler im Rahmen musealer (und organisatorischer) Vorgaben weitgehende Freiheiten besitzen, welche Inhalte sie tatsächlich bearbeiten.

Dies sei vereinfacht durch folgende Analogie veranschaulicht (Tabelle 4): Ein Diorama oder Schaukasten ist symbolisiert durch einen Rahmen, alle dort dargestellten Inhalte werden durch alle Kästchen im Inneren des Rahmens repräsentiert. Im Treatment S, durch weitgehende Selbststeuerung gekennzeichnet, z.B. durch Aufgaben wie „Erzähle, was Dich interessiert!“, ist nicht ein Inhalt (ein Kästchen) oder ein begrenzbarer Ausschnitt (wenige bestimmte Kästchen) das wahrscheinliche

Hauptobjekt der Bearbeitung der Schüler sondern es können alle Inhalte eines Dioramas oder Schaukastens in gleichem Maße sein (alle Kästchen). Das Treatment F dagegen beschränkt durch die Setzung von Multiple-Choice-Aufgaben den wahrscheinlichen Focus der Bearbeitung des Schülers auf wenige bestimmte Inhalte (wenige Kästchen). Die Mischform SF liegt mit einem mittleren Maß an Möglichkeiten und Bestimmtheit wahrscheinlicher Bearbeitung dazwischen (mittlere Menge an Kästchen). Diese Analogie gilt im Grunde für alle Stationen (vgl. Tabelle 1) wird aber am deutlichsten bei idealisierten Biotopen wie „Urwald im Fichtelgebirge“ (vgl. Unterrichtsliche Umsetzung: 3.5.2. Inhaltliche Analyse).

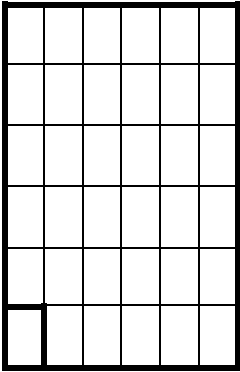
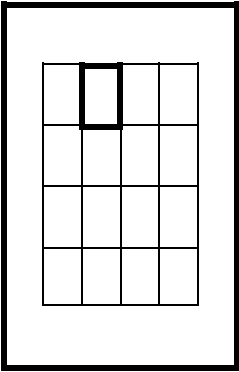
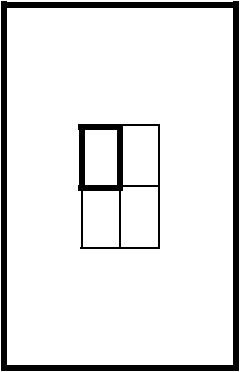
Tabelle 4: Veranschaulichung der wahrscheinlichen Treatmentinhalte. Erläuterungen im Text.

Aufgabe, z.B. „Erzähle, was Dich interessiert!“	Aufgabe, z.B. Vergleich (Text, Zeichnung)	Aufgabe, z.B. Multiple Choice
<p>S</p>  <p>Alle Inhalte Tiere, Pflanzen + ?</p>	<p>SF</p>  <p>Begrenztes Thema z.B. Spechte</p>	<p>F</p>  <p>Detail z.B. Schwarzspecht</p>

Diese unterschiedliche Vorhersagbarkeit bedingt die Möglichkeit einer relativen Steuerung der Resultate des Fragebogens und damit der Produktion beliebiger Artefakte. Durch folgendes Vorgehen wird versucht, diesen Effekt zu minimieren und größtmögliche Transparenz herzustellen.

Die Konstruktion des Fragebogens erfolgt in Hinblick auf den in Form unterschiedlicher Treatments gestalteten Besuch im Naturkundemuseum. Pro Treatment und bearbeitetes Diorama bzw. bearbeiteten Schaukastens gibt es ein Item, für drei Treatments und die sechs Arbeitsstationen, insgesamt 18 Items (*KogMit18*). Die im Folgenden erläuterte Vorgehensweise wird analog zur Darstellung der wahrscheinlichen Treatmentinhalte veranschaulicht (Tabelle 5). Ein F-Item muss dem F-Treatment entsprechen (Kästchen aus F). Ein SF-Item war inhaltlich Teil des SF-Treatments, aber nicht des F-Treatments (Kästchen aus SF, die keine Überschneidung mit Kästchen aus F bieten) und ein S-Item Teil des S-Treatments, aber nicht des F- und nicht des SF-Treatments (Kästchen aus S, die keine Überschneidung mit Kästchen aus SF und F bieten). Der Fragebogen besteht zu gleichen Teilen aus F-, SF- und S-Items. Er ist von allen Probanden unabhängig vom durchlaufenen Treatment zu beantworten.

Tabelle 5: Veranschaulichung der Itemwahl. Erläuterungen im Text.

S	SF	F
		
Aus Gesamtheit aller Inhalte z.B. Haselhuhn	Aus begrenztem Thema z.B. Stützenschwanz der Spechte	Aus Ansammlung von Details z.B. Kopfzeichnung des Schwarzspechts

Dies bedeutet für das resultierende Messinstrument (*KogMit18*) das Vorhandensein einer gleichgroßen Anzahl an Items, die besonders gut von S-, von SF- oder von F-Schülern, jeweils Probanden, die das entsprechende Treatment durchlaufen haben, bearbeitet werden können. Damit sollte jeder Schüler nach dem Museumsbesuch ein Spezialist für ein Drittel des Fragebogenteils *KogMit18* sein. In welchem Maße dies

für S-, SF- und F-Schüler gilt, soll an späterer Stelle diskutiert werden. Diese Treatmentbedingte Itemkonstruktion ist bei der Interpretation der Ergebnisse zu berücksichtigen.

Sieben weitere Items aus *KogMit25* behandeln methodische Fragen, z.B. den „Carry-Over-Effekt“, bzw. inhaltliche Nebenaspekte, z.B. das Kindermuseum (vgl. Anhang 20).

Im weiteren Bestreben um Transparenz werden die Anforderungen des Testteils *KogMit25* bezüglich der Inhalte des Museums präzise charakterisiert. Das erfolgt im folgenden Kapitel: Material und Methoden 4.2.2.

Ein zweites Problem könnte in der Ähnlichkeit der Formulierung von Items im Fragebogen und von Aufgaben im Treatment v. a. bei den F-Schülern liegen. In beiden Fällen bedient man sich Items mit Antwortvorgaben, so dass darin ein Vorteil für die Probanden des F-Treatments, für die Probanden des S-Treatments ein Nachteil zu sehen sein könnte.

Dem wird begegnet durch die Konstruktion eines weiteren Fragebogenteils, nämlich *KogO*, in dem Items ohne Antwortvorgaben verwendet werden und der durch seine sehr offene Itemformulierung eher den Aufgaben des S-Treatments entspricht. Ein Item lautet beispielsweise: „(...) Wie sieht für Dich der Urwald des Fichtelgebirges aus? Schreibe in Stichpunkten auf, was Dir dazu einfällt!“ Es gibt für jedes Diorama bzw. für jeden Schaukasten je ein ähnlich offen formuliertes, auf den Hauptinhalt der jeweiligen Station bezogenes Item. Ein zusätzlicher Vorteil des Testteils *KogO* liegt darin, dass diese Items, was den inhaltlichen Bezug zwischen Test und Treatment betrifft, infolge der Offenheit der Aufgabenstellung keinem der drei Treatments Vor- oder Nachteile bieten. Salopp formuliert: Jeder Schüler kann zeigen, was er weiß! *KogMit* mit treatmentbezogener Itemkonstruktion wird in *KogO* mit sehr offenen Items eine konträre Perspektive gegenübergestellt, die eine bessere Einschätzung der durch die drei Treatments vermittelten Lernergebnisse ermöglichen soll.

Bisher unbehandelt verbleibt die Frage nach der Validität dieser Items, nämlich ob sie wirklich das messen, was durch sie gemessen werden soll. Im kognitiven Bereich

des Fragebogens (*KogO* und *KogMit*) erscheint eine Bestimmung der Kriteriumsvalidität einfacher als im affektiven, da kognitive Leistungsmessung leichter zugänglich und ständiger Bestandteil des Schulalltags ist. Geht man davon aus, dass die kognitive Leistung eines Schülers als manifestes Merkmal (=Außenkriterium) gelten kann und man als Maß dafür die Noten eines Probanden im Schulunterricht heranziehen kann, dann lässt sich dieser Messwert, die Schulnote, mit den Ergebnissen zu den kognitiven Teilen des Fragebogens korrelieren. Die Korrelation zwischen der Halbjahresnote der Probanden in Biologie – bei 296 von 366 Probanden vorhanden – und den Testergebnissen wurde für die kognitiven Teile des Fragebogens berechnet: Für *KogMit* bezogen auf die Gesamtgruppe aller Schüler der Hauptuntersuchung ($N=296$) zeigen sich für die beiden Nachtests sehr geringe, wenngleich signifikante positive Korrelationen von $r(\text{NT I})=0,14$ und $r(\text{NT II})=0,16$. Ähnliches entdeckt man für *KogO*. Die Korrelationen sind etwas höher als bei *KogMit*: $r(\text{NT I})=0,23$ und $r(\text{NT II})=0,16$. Nach Zöfel (2002, S. 120) sind diese Werte als geringe bis sehr geringe Korrelationen einzuordnen. Dennoch lässt sich bezogen auf die kognitiven Teile des Fragebogens der schwache Zusammenhang formulieren: Je besser die Schulnote der Probanden in Biologie ist, desto besser sind die Ergebnisse im Fragebogen. Eine Zusammenschau aller Werte findet sich im Anhang 21. Schwächen dieser Vorgehensweise sind darin zu sehen, dass Noten als hinreichend valides und reliables Merkmal für kognitive Leitungsfähigkeit kritisch zu sehen sind, im Besonderen die Validität dieses Außenkriteriums anzuzweifeln ist. Möglicherweise könnte sich dies gerade darin ausdrücken, dass nach schulischer Definition schlechte Schüler in diesen kognitiven Tests von ihrer üblichen Schulleistung abweichende Ergebnisse haben. Dies könnte fehlende oder geringe Korrelationen bedingen. Im Kern liegen hier die Schwierigkeiten einer Bestimmung der Kriteriumsvalidität auf einer den affektiven Fragebogenteilen vergleichbaren Ebene.

4.2.2. Inhaltliche Analyse des Fragebogens

Zuerst sei die affektive Ebene betrachtet. Bewertet wird das „Gefallen“ des Besuches (*AffMu*, erstes Item), der Arbeitsstationen (*AffMu*, übrige Items) und von Einzelelementen eines Dioramas und eines Schaukastens (*AffMuUr*, *AffMuPla*). Es soll ferner gezeichnet werden, was den Schülern am besten gefallen hat (*Freie Äußerung*, letztes Item), und die Absicht über einen weiteren Besuch geäußert werden (*Freie Äuße-*

runge, erstes Item). Zentrales Merkmal ist Klarheit und Einfachheit. Besonders bei den affektiven Items wird auf zu große sprachliche und inhaltliche Differenzierung zu Gunsten von möglichst guter Verständlichkeit verzichtet. Beispielsweise lautet die Frage nicht „Was hast Du empfunden, während Du die Aufgaben zum Diorama ‚Wacholderheide im Frankenjura‘ bearbeitet hast und wie hat Dich das Thema des Dioramas berührt?“ sondern „Wie hat Dir die Station ‚Wacholderheide im Frankenjura‘ gefallen?“ Obwohl die erste Version wahrscheinlich inhaltlich und sprachlich den Erlebnischarakter der Bearbeitung und die den Inhalt betreffenden Emotionen des Probanden exakter treffen würde, wird im Bestreben um Verständlichkeit und Akzeptanz bei den Schülern die zweite gewählt. Inhaltliche Schwierigkeiten der Probanden mit den Aufgabenstellungen sind nicht zu erwarten. Probleme bezüglich der Lese- und der Schreibfähigkeiten, kann man für diese Probanden, bei denen es sich ausschließlich um Gymnasiasten handelt, ausschließen. Verständnisprobleme werden durch weitgehende Vermeidung komplexer, ungewöhnlicher oder aus sonstigen Gründen schwieriger sprachlicher Konstruktionen oder Wortwahl möglichst ausgeschlossen. Das gilt für die Verwendung von Fachtermini nur begrenzt, wobei auch hier den Schülern das Verständnis so weit wie möglich erleichtert wird. Beispielsweise werden die Stationen des Museums, wann immer sich ein Item mit einer bestimmten Station beschäftigt, nicht ausschließlich durch den Stationentitel, z.B. „Sandgrube in Nordbayern“, sondern zusätzlich durch ein Symbol, das an der Station und auf allen Aufgaben zu dieser Station zu finden ist, im Falle der „Sandgrube“ ein Froschlurch, veranschaulicht (vgl. Abbildung 38). Kurz: Schwierigkeiten Inhalte richtig zu erfassen, werden so weit wie möglich vermieden. Inwiefern Ebenen der durch den Fragebogen vermittelten Kommunikation jenseits der kognitiven Verständlichkeit beachtet werden, soll im Rahmen der Beschreibung des Versuchsablaufs (siehe unten Material und Methoden 7.) geschildert werden.



Abbildung 38: Rechts oben das Symbol für die Sandgrube, ein Froschlurch.

Für *KogO*, entwickelt im Blick auf eine Perspektive der kognitiven Einschätzung der Schüler, gelten ebenfalls die Grundsätze von Einfachheit und Verständlichkeit. Die zentrale Frage zielt darauf ab, den Schülern Raum zu geben für das, was sie verstehen bzw. sich unter dem erfragten Inhalt vorstellen. Für ein Biotop lautet beispielsweise die Frage: „Wie sieht für Dich der Urwald im Fichtelgebirge aus?“; und die Arbeitsanweisung heißt: „Schreibe in Stichpunkten auf, was Dir dazu einfällt!“ Es wird bewusst auf Einschränkungen oder Präzisierungen verzichtet, um dem Schüler Raum für eine betont subjektive Sicht zu geben. Die Items aus *KogO* bzw. Teile dieser Items erinnern z.T. an Assoziationsfragen. Entsprechend kann man kognitive, affektive und nicht fachlich einzuordnende Schüleraussagen erwarten. An dieser Stelle ist es wichtig, sich zu vergegenwärtigen, dass die Art des Items das Resultat maßgeblich bestimmen kann. Bei der Konstruktion dieser Items spielte die Theorie des Konstruktivismus als eine diese Untersuchung begründende Idee eine wichtige Rolle, denn dieser Fragebogenteil soll eine Perspektive jenseits der rein instruktionalen Sicht des Wissenserwerbs, wie sie im späteren Fragebogenteil *KogMit* dominiert, ermöglichen. Bei der Auswertung wird diese sehr offene Art der Items berücksichtigt. Jedoch weniger als sieben Prozent der Schülerantworten berühren u.a. affektive Ebenen. Nicht sinnvoll fachlich einzuordnende Aussagen machen nur ca. 1,2% der Schülerantworten aus. Eine v. a. kognitiv, fachlich ausgerichtete Auswertung entspricht durchaus den Schülerantworten.

Eine ganz andere Art, kognitives Wissen zu bestimmen, ist im Fragebogenteil *KogMit* verwirklicht und soll hier, wie oben angedeutet, präzise analysiert werden. Das Ziel dieser Analyse ist eine möglichst genaue Charakterisierung der Anforderung der Items. Man kann sich folgende „ideale“ Situation vorstellen: Die Probanden wissen vor dem Unterricht (hier vor dem Treatment) nichts und werden allein durch das Treatment bzw. die Treatments „befähigt“, bestimmte Aufgaben (Items der Fragebögen) lösen zu können. Dafür sind unterschiedliche Ebenen kognitiver Operationen zu meistern, die sich didaktisch beschreiben lassen. Die vorgefundene Situation ist von diesem beschriebenen Ideal nicht sehr weit entfernt. Wie Experteneinschätzungen belegen, können Gymnasiasten der fünften Jahrgangsstufe vor dem Treatment wahrscheinlich nur einen sehr kleinen Teil der Items richtig bearbeiten (vgl. Material und Methoden 2. Vorstudien). Nicht ideal ist das zu erwartende nicht-homogene Niveau dieses (geringen) Vorwissens. Die 366 Probanden sind aus 14 Schulklassen,

die sich erst vor Monaten aus einer noch größeren Anzahl von Grundschulklassen zusammengefunden haben. Selbst die in Bayern noch recht deutlichen Vorgaben des Grundschullehrplans von 1981 (KMBI I So.-Nr. 20/1981), der zum Zeitpunkt des Grundschulbesuchs der Probanden Gültigkeit hatte, erlauben keine eindeutigen Vorannahmen.

Das erste relativ einfache didaktische Verfahren zur Kategorisierung von Aufgaben orientiert sich an den Vorgaben des Deutschen Bildungsrates (1970) und wird in der aktuellen Schulpraxis – zumindest im Bundesland Bayern – zur Einschätzung von Aufgaben in Schulprüfungen, somit von Items in Testverfahren, herangezogen. Ein komplexeres Verfahren richtet sich nach der überarbeiteten Form (Anderson et al. 2001) der sehr ausführlichen und differenzierten Bloom'schen Taxonomie, die erstmals vollständig im Jahre 1956 publiziert wurde (Bloom 1956, zitiert nach Anderson & Sosniak 1994).

Die erste Art einer objektivierenden Beschreibung der kognitiven Anforderungen besteht in den vom Deutschen Bildungsrat (1970) vorgeschlagenen Stufen von Unterrichtszielen, kurz zusammengefasst von Eschenhagen et al. (2001; S. 180):

1. Reproduktion: Wiedergabe von Sachverhalten aus dem Gedächtnis.
2. Reorganisation: selbständige Neuordnung bekannter Sachverhalte zu einer neuen, komplexen Struktur.
3. Transfer: Übertragen von bekannten Zusammenhängen auf eine Struktur neuer Sachverhalte.
4. Problemlösen: Lösen neuartiger Aufgaben bzw. Finden neuartiger Erklärungen für bekannte Sachverhalte; konstruktive Kritik bekannter Lösungsvorschläge.

Eine Beurteilung, zu welcher Stufe eine kognitive Schülerleistung zu zählen ist, kann nur durch genaue Kenntnis des Unterrichts erfolgen. Die drei Treatments, bzw. deren Rezeption durch die Schüler, bedingen unterschiedliche Nuancen für die Kategorisierung. Darum können die Einteilungen in die jeweiligen Stufen nur ungefähr erfolgen. Sie beziehen sich auf den Nachtest I. Eine Sicherung der Objektivität durch Bestimmung der Signierobjektivität mit Hilfe eines zweiten Beurteilers ist nicht möglich, da in diesem Fall zu genaue Kenntnisse der Treatments und des Museums nötig sind.

Statt dessen werden die Urteile mit einem zweiten Experten (Gymnasiallehrer) im Gespräch abgestimmt.

Tabelle 6: Übersicht über die Einschätzung der kognitiven Stufe der Anforderung der Items im Fragebogenteil *KogMit18*. Aufschlüsselung für die einzelnen Items: siehe Anhang 22.

Expertenbewertung	S-Items	SF-Items	F-Items
Reproduktion	5	1	4
Reorganisation	-	1	-
Transfer	1	3	2
Problemlösung	-	1	-

Das Ergebnis der Kategorisierung ist heterogen, das Anforderungsniveau z. T. niedrig. Während bei S- und F-Items v. a. Reproduktion verlangt wird, erfordern SF-Items höhere kognitive Leistungen. Dies sollte bei späterer Analyse der Ergebnisse beachtet werden.

Zur Bloom'schen Taxonomie (Anderson et al. 2001, S. 27 ff.; vgl. Anderson & Sosniak 1994 und Bloom 1976) angewandt auf den Fragebogenteil *KogMit25*: Dieses System ist so differenziert und umfassend, dass eine kurze Vorstellung nötig erscheint. Die englischen Fachbegriffe und Erläuterungen werden, um Verfälschungen zu vermeiden, nicht übersetzt. Man unterscheidet vier „knowledge dimensions“ und sechs „cognitive process dimensions“. Damit sind kognitive Anforderungen in einer Matrix eindeutig zu verorten (Zeilen: „knowledge dimensions“, Spalten: „cognitive process dimensions“). Beide Auflistungen folgen dem Ordnungsprinzip aufsteigender Komplexität.

Die „knowledge dimensions“ sind:

- „factual knowledge“ (F): „discrete, isolated elements“, „bits of information“, e.g. „terminology, specific details and elements“.
- „conceptual knowledge“ (C): „more complex, organized knowledge forms“, e.g. „classifications and categories“, „principles and generalizations“, etc.

- “procedural knowledge”(P): “knowledge of how to do something”, e.g. “specific skills”, “techniques and methods”, “criteria (when to do what)”
- “Metakognitive knowledge” (M): “knowledge about cognition in general as well as the awareness of and knowledge about ones own cognition”, e.g. “strategic knowledge”, “self knowledge”, etc.

Die „cognitive process dimensions“ lassen sich jeweils durch bestimmte Verben, die ebenfalls genannt werden, differenziert verorten:

- 1. Remember: a) Recognizing
 b) Recalling
- 2. Understand: a) Interpreting
 b) Exemplifying
 c) Classifying
 d) Summarizing
 e) Inferring
 f) Comparing
 g) Explaining
- 3. Apply: a) Executing
 b) Implementing
- 4. Analyze: a) Differentiating
 b) Organizing
 c) Attributing
- 5. Evaluate: a) Checking
 b) Critiquing
- 6. Create: a) Generating
 b) Planning
 c) Producing

Zur Analyse werden die Items zunächst in die Einzelelemente der Anforderungen zerlegt. Diese sind wie Feinlernziele operationalisierbar verfasst (vgl. Möller 1976, S. 82), so dass man gedanklich vor jeden Einzelausdruck die Formulierung „Die Schüler sollen...“ setzen kann. Diesen Präzisierungen sind die entsprechenden Bloom'schen Kategorisierungen, „knowledge dimension“ und „cognitive process“, zugeordnet und entsprechend der Bloom'schen taxonomischen Matrix eingeordnet.

Es werden zwei Ansätze zur Quantifizierung der Anforderungsstufen versucht. Die erste Möglichkeit ist sehr transparent und unkompliziert. Sie besteht darin, die Einzelanforderungen zu zählen (Tabelle 7: B1). Dabei wird angenommen, je mehr Elemente ein Item umfasst, desto komplexer ist die Anforderung. Beispielsweise verlangt die richtige Beurteilung der Behauptung „Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.“ aus dem ersten Item die Anforderungen „terminology recognizing“ und „specific detail recalling“. Damit erforderte die Lösung des Items zwei „Einzelanforderungen“. Die zweite Möglichkeit liegt darin, unter Zuhilfenahme der Bloom'schen Kategorisierungen die Anforderungen zu gewichten. Das geschieht, indem die Anforderungen mit dem Faktor eins, zwei oder drei multipliziert werden. Die gewichteten Werte addieren sich zum Gesamtwert des Items (Gewichtung: siehe Anhang 23). Betrachtet man wiederum das erste Item, so lassen sich zunächst auf der Ebene der „knowledge dimensions“ „terminology“ und „specific detail“ unterscheiden. Hier wurde die Kategorie „terminology“ geringer eingeschätzt, darum mit dem Faktor eins belegt, als „specific detail“, das mit Faktor zwei eingeschätzt wurde. Auf der Ebene „cognitive process dimensions“ wurde „recognizing“ geringer eingeschätzt (Faktor eins) als „recalling“ (Faktor zwei). Damit resultiert daraus für das erste Item: [„terminology“] x 1 + [„recognizing“] x 1 + [„specific detail“] x 2 + [„recalling“] x 2 = 1x1 + 1x1 + 1x2 + 1x2 = 6. So wird das erste Item in der Summe mit sechs Punkten bewertet. Dieses Verfahren erlaubt, postulierte unterschiedliche kognitive Ebenen der Anforderungen weitgehend gemäß der Bloom'schen Hierarchisierung zu berücksichtigen (Tabelle 7: B2). Bei beiden Arten des Vorgehens ist zu bedenken, dass der gewonnene Wert den Charakter einer Mischbewertung hat und zwar zwischen der Anforderung, die man der Dimension des zu lernenden Inhalts beimisst, und der Anforderung der Dimension des jeweiligen kognitiven Tuns. Beide

Verfahren beziehen mögliches Vorwissen der Probanden aus Schulunterricht und Alltag nicht ein.

Das Verfahren dieser Bloom'schen Kategorisierungen lässt sich für den Fragebogen-*KogMit18* sehr gut anwenden. Für die „knowledge dimension“ dominiert klar „factual knowledge“ vor „conceptual knowledge“, andere Dimensionen treten nicht auf, für die „cognitive process dimension“ überwiegt „remember“ vor „understand“ und „apply“, andere Dimensionen treten nicht auf.

Die Quantifizierung der Anforderungsstufen durch die Bestimmung der Summe der Einzelanforderungen (B1) erbringt für einzelne Items Werte von zwei bis sechs, in der Übersicht insgesamt für S-Items 18, für F-Items 24 und für SF-Items 27 Einzel-elemente. Der zweite Ansatz mit Gewichtung der jeweiligen kognitiven Anforderungsebene (B2) liefert für Einzelitems Werte zwischen vier und 18. Das sind erhebliche Unterschiede. Resultat des Verfahrens ist jedoch die gleiche Reihung: S-Items 54, für F-Items 72 und SF-Items 85. Auch die relativen Unterschiede zwischen den Itemtypen bleiben bei diesen beiden Verfahren in etwa gleich.

Die Quantifizierungen nach der Bloom'schen Kategorisierung von *KogMit18* stimmen in der resultierenden Reihung nach Anforderungsgrad der Items mit „S < F < SF“ mit der Analyse nach dem Deutschen Bildungsrat (1970) überein. Bei beiden Verfahren überwiegen Reproduktion bzw. die „cognitive process dimension“ „remember“. Der Aussage von Bortz & Döring (1995, S. 196), durch Items mit Alternativantworten seien ausschließlich einfache Reproduktionsaufgaben zu testen, ist an dieser Stelle ausdrücklich zu widersprechen. Wie die Analysen zum kognitiven Anforderungsniveau zeigen, kann die Ebene der Reproduktion deutlich überstiegen werden (vgl. Tabelle 5 und 6 bzw. Anhang 22 und 24).

Tabelle 7: Inhaltliche Analyse des Fragebogenteils *KogMit25* mit Hilfe der Bloom'schen Taxonomie (Anderson et al. 2001, S. 27 ff.) und zwei Versuche zur Bewertung des kognitiven Anforderungsniveaus (B1, B2). Weitere Erklärungen finden sich im Text. Dargestellt sind hier die Items zur Station Urwald. Die Kürzel in der Spalte „Matrix“, z.B. F1a, beziehen sich auf die bei der Erläuterung der Bloom'schen Taxonomie eingeführten Abkürzungen und Gliederungspunkte. Tr.=Treatment, B1=Beurteilungswert 1, B2=Beurteilungswert 2. Für die vollständige Analyse siehe Anhang 24.

Station	Tr.	Items / Einzelanforderungen	B1		B2	Matrix
Urwald	S	Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.	$\Sigma 2$		$\Sigma 6$	
		Artbezeichnung wieder erkennen können		terminology	1	F1a
		Merkmale (Zeichnung, Schnabel) der Waldschnepfe vorstellen können ⁹		recognizing	1	
				specific detail	2	F1b
				recalling	2	
	SF	Buntspechte sind kleiner als Grünspechte.	$\Sigma 5$		$\Sigma 13$	
		Artbezeichnungen wieder erkennen können (2x)		terminology	1x2	F1a
		diese Vögel in der richtigen Größe vorstellen können (2x)		recognizing	1x2	
		Größe vergleichen können		detail	1x2	F1b
				recalling	2x2	
				details	-	F2f
				comparing	3	
	F	Bild (Brachvogelschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	$\Sigma 2$		$\Sigma 5$	
		Artbezeichnung wieder erkennen können		terminology	1	F1a
		Schnabel eines Schwarzspechtes wieder erkennen können		recognizing	1	
				specific detail	2	F1a
				recognizing	1	

⁹ Das Verb „vorstellen“ ist hier verwendet im Sinne von „sich etwas bildhaft ins Gedächtnis (zurück)rufen“, nicht im Sinne von „sich etwas ausdenken, träumen, phantasieren“.

4.3. Untersuchungstechnische Betrachtung

4.3.1. Untersuchungstechnische Analyse des Fragebogens

Für *KogMit18* (vgl. Material und Methoden 4.2.1.), im Sinne von Lienert (1969, S. 21 f.) aufzufassen als Leistungstest, wird eine Reliabilitätsanalyse (Zöfel 2002, S. 231 ff.) (=Itemanalyse, Bortz & Döring 1995, S. 198 ff.) durchgeführt. Diese 18 Items stellen die maßgebliche Grundlage für Aussagen zu Wissensstand und –veränderungen dar. Darum erscheinen die folgenden aufwendigen Überlegungen gerechtfertigt.

Der Reliabilitätsanalyse lassen sich unterschiedliche Stichproben zu Grunde legen. Nahe liegend ist, jedes Item auf die Gesamtheit aller Schüler zu beziehen, die das entsprechende Item sinnvoll auswertbar bearbeitet haben, und zu differenzieren zwischen „Item richtig“ und „Item Rest“, entsprechend kodiert mit „1“ und „0“. Dabei bestünde „Item Rest“ aus falschen Antworten und Antworten der Kategorie „weiß nicht“. Ein Zusammenfassen von falsch beantworteten Items und „weiß nicht“, erscheint plausibel, da sich Schüler, die ein Item wahrscheinlich nicht lösen können, sicher nicht selten für die Antwort „weiß nicht“ entscheiden.

Da es sich hier um ein grundsätzliches Problem handelt, soll es genauer betrachtet werden. Analysiert man das Verfahren im Einzelnen so wird klar, dass es sich für den Schüler um Entscheidungen auf unterschiedlicher Ebene handelt. Die erste Entscheidung besteht in der Verweigerung einer Antwort oder der Entscheidung, sich sinnvoll im Fragebogen zu äußern. Dabei wirft diese Ebene das Problem auf, ob es sich bei sinnloser Bearbeitung tatsächlich um eine Entscheidung oder um ein Missverständnis handelt. Das ist nicht zu klären. Auf den konkreten Fall bezogen, kann man nicht ausschließen, dass sich Schüler, die das Item inhaltlich nicht beantworten können, besonders häufig für die Nicht-Beantwortung entscheiden. Die zweite Ebene der Entscheidung liegt im Rahmen eines auswertbar bearbeiteten Fragebogens, nämlich ob ein Schüler die inhaltliche Frage für lösbar hält oder nicht. Er wird daraufhin „weiß nicht“ oder eine der Antwortmöglichkeiten ankreuzen. Erst auf der dritten Ebene liegt die Entscheidung für die richtige oder die falsche Antwort.

Diese Analyse zeigt auf, an welchem Punkt die oben als „nahe liegend“ bezeichnete Betrachtung („Item richtig“ vs. „Item Rest“) ansetzt, nämlich auf der zweiten Ebene der Analyse. Die insgesamt geringe Anzahl von nicht regelgerecht bearbeiteten Items, lässt das Verfahren, diese Daten als fehlend zu betrachten, ohne die präzisen Ursachen für das Fehlen weiter zu verfolgen, angemessen erscheinen. Lienert (1961, S. 97) schlägt zur Lösung des Problems fehlender Daten ebenfalls dieses Verfahren vor. Als weitere Möglichkeit für die Reliabilitätsanalyse bietet sich die dritte Ebene an, d.h. die Stichprobe besteht aus Schülern, die sich bewusst für eine Antwort im Fragebogen entscheiden („Item richtig“ vs. „Item falsch“). Die Schüler, die die Kategorie „weiß nicht“ wählen, sich also bewusst nicht entscheiden, bleiben unberücksichtigt. Für diese beiden Möglichkeiten wird der Reliabilitätskoeffizient „Cronbachs Alpha“ (Bortz & Döring 1995, S. 184) zu den drei Testzeitpunkten bestimmt (vgl. Tabelle 8).

Tabelle 8: Reliabilitäten zum Testteil *KogMit18* .

	Vortest	Nachtest I	Nachtest II
„Item richtig“ vs. „Item Rest“	$\alpha=0,66$	$\alpha=0,59$	$\alpha=0,70$
„Item richtig“ vs. „Item falsch“	$\alpha=0,61$	$\alpha=0,68$	$\alpha=0,76$

Die Koeffizienten für beide Möglichkeiten sind nicht sehr hoch, aber zur Differenzierung von Gruppen akzeptabel. Die von Lienert (1969, S. 246) angegebene Grenze, jenseits derer ein Test für eine sinnvolle Differenzierung von Einzelindividuen nicht mehr herangezogen werden sollte, liegt bei einem Reliabilitätskoeffizienten von unter 0,6, für eine Differenzierung von Gruppen liegt der Wert niedriger. Damit sind die vorliegenden Reliabilitäten, die ja ausschließlich zur Differenzierung von Gruppen verwendet werden, völlig ausreichend. Eine Erhöhung eines Reliabilitätskoeffizienten von beispielsweise 0,7 auf 0,9 bei 20 Items wird in Lienert (1969, S. 244) exemplarisch berechnet. Dies führt zu einer ungefähren Erhöhung der Itemanzahl um das Vierfache, somit auf 80 Items. Die Itemanzahl von *KogMit* liegt etwa in der Größenordnung des Ausgangswertes von Lienert. Eine entsprechende Erhöhung der Reliabilität kommt nicht in Frage, denn damit wäre der Test nicht mehr „ökonomisch“. Eine Anzahl von über 70 Items nur in diesem einen Teil des Fragebogens, würde dazu führen, dass der gesamte Test von den Schülern nicht mehr in vertretbarer Zeit zu bearbeiten wäre.

Für die nachfolgenden Berechnungen im Rahmen der Reliabilitätsanalyse soll aus folgenden Gründen die erste Möglichkeit („Item richtig“ vs. „Item Rest“) weiterhin Verwendung finden. Bei sonst in etwa gleicher Weise plausiblen Argumenten gehen zu allen Testzeitpunkten bei allen Items über 80% aller Probanden (Listenweise Fallausschluss) der Hauptuntersuchung in die Berechnungen ein. Bei der zweiten Möglichkeit („Item richtig“ vs. „Item falsch“) sind es z. T. weniger als 20 %, nämlich bei hohen Anteilen von Schülern, die „weiß nicht“ angeben. Damit wären die Angaben zu Reliabilität, Trennschärfe und Schwierigkeitsgrad nicht für die gesamte Stichprobe von 366 Schülern gültig, sondern jeweils nur für die Schüler, die sich für eine Beantwortung der Items entscheiden. Folglich wäre es erforderlich, in nachfolgenden Berechnungen diesen Punkt zu berücksichtigen. Dies wird als zu aufwendig eingeschätzt.

Die Trennschärfe eines Items sagt aus, inwieweit dieses Item repräsentativ für den gesamten Fragebogen ist (Bortz & Döring 1995, S. 199 f.). Berechnet wird die korrigierte Trennschärfe, d.h. die Beantwortung eines Items wird korreliert mit dem Gesamtergebnis des Tests, wobei das aktuelle Item für die Bestimmung des Gesamtwertes unberücksichtigt bleibt (Zöfel 2002, S. 235). Alle Trennschärfekoeffizienten liegen im positiven Bereich, d.h. „gute“ Probanden beantworten das entsprechende Item mit größerer Wahrscheinlichkeit korrekt als „schlechte“. Bezüglich der drei Testzeitpunkte liegen bei Weitem die meisten Werte im Bereich von 0,2 bis 0,4, damit in einem niedrigen bis mittleren Bereich (Bortz & Döring 1995, S. 200; vgl. Anhang 25).

Der Schwierigkeitsgrad oder Schwierigkeitsindex gibt an, wie viel Prozent der Probanden ein Item richtig lösen können (Leibold 1997, S. 95f, Zöfel 2002, S. 234 f.). Dabei gilt: Je höher der Schwierigkeitsindex ist, desto leichter ist ein Item. Nach Leibold (1997, S. 95) sind Schwierigkeitsgrade unter 10% und über 90% als zu schwer bzw. zu leicht abzulehnen. Indices von 10 bis 20% und 80 bis 90% gelten als akzeptabel, die Aufgaben aber als schwer bzw. leicht. Schwierigkeitsgrade von 20 bis 80% gelten als angemessen. Alle Items liegen innerhalb der Grenzen der obigen statistischen Konvention (vgl. Anhang 25).

Ziel ist, den Probanden im Fragebogen nicht ausschließlich Items mittleren Schwierigkeitsgrades zu bieten, sondern eine Mischung aus eher schweren, eher leichten und mittleren Items. Naturgemäß fallen im Vortest mehr Items eher schwer, in Nachtests mehr Items eher leicht aus. Dabei sind „Trennschärfekoeffizient und Schwierigkeitsindex (...) nicht unabhängig voneinander“ (Zöfel 2002, S. 236). Für Items mittlerer Schwierigkeit ist die Trennschärfe am höchsten, für schwere und leichte Items fällt sie jeweils ab.

Weiter wird für die Testteile *KogO*, *AffMu* und „*Tieraffektivität*“ die Reliabilität bestimmt. Wie Tabelle 9 belegt sind die Reliabilitätskoeffizienten der Nachtests in allen Fällen akzeptabel.

Tabelle 9: Reliabilitätskoeffizienten: Cronbachs α . Ausgeschlossen werden bei *AffMu* die Schüler, die angeben, sich nicht mehr an den in der Frage angesprochenen Inhalt erinnern zu können. Bei „*Tieraffektivität*“ werden die Schüler ausgeschlossen, die das entsprechende Tier nicht kennen und die Schüler, die angeben, das fiktive „Kontrolltier“ (Sumpfdotter Schnepfe) zu kennen, denn ihre Angaben sind diesen Testteil betreffend offensichtlich nicht valide.

	VT	NT I	NT II
<i>KogO</i>	0,48	0,69	0,74
N	366	366	366
<i>AffMu</i>	-	0,75	0,85
N	-	316	303
„ <i>Tieraffektivität</i> “	0,77	0,85	0,82
N	233	160	112

4.3.2. Auswertungsobjektivität

Bei den Items mit Antwortvorgabe, bei denen zuvor Kodierung, Auswertung und Interpretation standardisiert festgelegt werden, ist eine Überprüfung der Auswertungsobjektivität hinfällig. Die möglichen Antworten der Schüler sind vorgegeben. Spielraum des Begutachters bei der Bewertung der Antworten existiert nicht. Dieses quantitative Verfahren gewährt hinreichende Auswertungsobjektivität (Bortz & Döring 1995, S. 180 f.).

Bei der Auswertung von Schüleräußerungen zu Items ohne Antwortvorgaben, die hier aus Zeichnungen und kurzen Texten bestehen, besteht der hauptsächliche Signierungsaspekt im Inhalt der verbalen oder zeichnerischen Äußerung der Schüler (Rost 1996, S. 84 ff.). Die Kategorisierungen sind z. T. nicht so weit objektivierbar, dass von einer völligen Übereinstimmung unterschiedlicher Begutachter auszugehen ist. Man muss differenzieren zwischen quantitativen und qualitativen Daten. Während beim quantitativen Ansatz ein Ausschnitt der Beobachtungsrealität gemessen wird und diese Werte in der Folge statistisch verarbeitet werden, ist es beim qualitativen Ansatz erforderlich, Verbalisierungen oder Zeichnungen der Erfahrungswirklichkeit der Probanden zunächst zu interpretieren (Bortz & Döring 1995, S. 271 ff.). Als quantitativ zu behandelnde Items ohne Antwortvorgabe werden Schülerantworten gesehen, die sich auf die Nennung unstrittiger Fakten beziehen, z.B. die Aufzählung von Details aus Dioramen oder die Nennung von Merkmalen z.B. eines Archaeopteryx. In diesen Fällen erübrigt es sich, so wie bei den Items mit Antwortvorgaben, die Auswertungsobjektivität zu überprüfen.

Offene bei der Auswertung interpretationsbedürftige Items werden überprüft. Dazu zählen z.B. zeichnerische Äußerungen der Schüler. Aus diesen Items, verbal oder zeichnerisch bearbeitet, werden randomisiert Stichproben gegebener Antworten ausgewählt und es wird ein Übereinstimmungskoeffizient, nämlich Cohens Kappa-Koeffizient (κ), bestimmt, der den Grad der Übereinstimmung nominal skalierten Daten zweier unabhängiger Beurteiler als Wert zwischen 0 und 1 angibt (Cohen 1968). κ ist dem Übereinstimmungsmaß als *Anteil der übereinstimmenden Beurteilungen an der Gesamtanzahl der Vergleiche* vorzuziehen, da bei diesem zweiten Übereinstimmungsmaß zufällige Übereinstimmung zwischen den Beurteilern nicht berücksichtigt wird (Cohen 1968, Zöfel 2002, S. 167 ff.). Zweiter Begutachter ist eine Person, die an der Untersuchung in keiner Form beteiligt ist, den Lindenhof jedoch kennt. Fragebögen, in denen Schüler sich zu einem der unten genannten Items nicht äußern, werden nicht in die Berechnung von κ miteinbezogen, da eine Nicht-Antwort keinen subjektiven Bewertungsspielraum zulässt, so dass durch eine Mitverwendung nicht-gegebener Antworten κ künstlich erhöht würde. Zur Einschätzung der Auswertungsobjektivität wurde bei $N > 250$ eine Stichprobengröße von 10 % der gegebenen Antworten für ausreichend erachtet, bei deutlich geringerer Stichprobengröße ($N=88$)

wurde ein Vielfaches dieses Anteils (40%) zur Bestimmung von κ herangezogen. „ κ -Werte über 0,75 zeigen eine (sehr) hohe Urteilerkonkordanz an, Werte zwischen 0,40 und 0,75 stehen für mäßige bis gute Übereinstimmung, (...) Werte unter 0,40 [zeigen] eine schlechte (...) Übereinstimmung zwischen den Beurteilern [an]“ (Diehl & Staufenbiel 2002, S. 161 f.). Nach Bortz & Döring (1995, S. 254) steht ein κ -Wert von über 0,70 für gute Beurteilerübereinstimmung, also akzeptable Signierobjektivität (vgl. Tabelle 10).

Tabelle 10: Cohens Kappa-Koeffizient (κ) für Items ohne Antwortvorgaben, die subjektiven Beurteilerspielraum zulassen.

Items	Bewertungskategorien	Cohens κ
Vortest (Soziodemographisch): Im Anschluss an die Frage, ob der Lindenhof bekannt ist: „Wenn ja, was hast Du dort gemacht?“	<ul style="list-style-type: none"> • Kinderkurs (incl. Kindergeburtstag) • Biologischer Kurs • Ausstellung (incl. Museen) • Mehrfachbesuch • Sonstiges <u>nominales Skalenniveau</u>	$\kappa=0,834$ (aus 40%; N=88 im VT)
Nachtest I und II (KogO): „Wie hängen das Gewicht von Vögeln und ihr Platzbedarf zusammen? Antworte in einem Satz!“	<ul style="list-style-type: none"> • Völlig falsch • korrekte(s) Detail(s) ohne Hauptinhalt wiederzugeben • Hauptinhalt exemplarisch, aber nicht abstrahiert • Hauptinhalt verallgemeinert dargestellt mit sprachlichen Unzulässigkeiten • Hauptinhalt in verallgemeinerter Form inhaltlich und sprachlich richtig präsentiert <u>ordinales Skalenniveau</u>	[$\kappa=0,793$ (aus 10%; N= 265 im NT I)] $r=0,828$ (Rangkorrelationskoeffizient nach Spearman) ¹⁰

¹⁰ Im Fall dieses Items ist wegen der ordinalskalierten Variablen eher der Rangkorrelationskoeffizient nach Spearman zu berechnen (Zöfel 2002, S. 88f, 124 und 169). Der Wert $r=0,828$ gibt eine hohe Korrelation an, spricht demnach für gute Beurteilerübereinstimmung (Zöfel 2002, S. 120).

Nachtest I und II (<i>Freie Äußerungen</i>): „ <i>Willst Du zu (dem) (...) Museumsbesuch noch etwas sagen?</i> “	<ul style="list-style-type: none"> • Positive Bewertung • negative Bewertung • positive und negative Bewertung • kritische Anmerkung bzw. Anregung; <u>nominales Skalenniveau</u>	$\kappa=0,675$ (aus 10%; N= 256 im NT I)
Nachtest I und II (<i>Freie Äußerungen</i>): „ <i>Wenn Du noch Zeit hast, dann zeichne auf die Rückseite dieses Blattes, was Dir am Lindenhof am allerbesten gefallen hat.</i> “	<ul style="list-style-type: none"> • Inhalte des Kindermuseums • Inhalte des Naturkundemuseums • gleichzeitig Inhalte des Naturkunde- und des Kindermuseums • Inhalt, der nichts mit dem Lindenhof zu tun hat • Zwergmäuse • Betreuerin • nicht erkennbar; <u>nominales Skalenniveau</u>	$\kappa=0,788$ (aus 10%; N= 291 im NT I)

Wie diese Kenngrößen zur Beurteilerübereinstimmung zeigen, sind die obigen Items wie erwartet nicht völlig objektivierbar. Dennoch liegt nur in einem Fall κ mit 0,675 knapp unter der Grenze akzeptabler Signierobjektivität von Bortz & Döring, nach Diehl & Staufenbiel dagegen etwa im Bereich guter Übereinstimmung (vgl. oben). Aufgrund dieser Analysen wird für die Fragebögen von Vortest, Nachtest I und II ausreichende, z. T. gute bis sehr gute Auswertungsobjektivität angenommen.

4.3.3. Testverfälschungen

Testverfälschungen können absichtlich oder unabsichtlich von den Probanden herbeigeführt werden (Bortz & Döring 1995, S. 210 ff.). Ein Beispiel für absichtliche Verfälschung ist, Wissen durch Raten zu simulieren oder Nichtwissen durch „Dummstellen“. Unabsichtliche Verfälschungen treten z.B. durch Konzentrationsmängel oder Unaufmerksamkeiten auf.

Die Reihenfolge der Elemente des Fragebogens ist für diesen Punkt bedeutsam (vgl. Tabelle 3 und Anhang 8 bis 15). Sie folgt der Prämisse, das Messergebnis eines Testteils nicht durch andere Items des Fragebogens zu beeinflussen. Exemplarisch sei dies für den Vortest erläutert. Zuerst werden soziodemographische Angaben erbeten, einfache Auskünfte, die den Einstieg zur Bearbeitung des Tests erleichtern sollen. Es folgen Items zur affektiven Einschätzung vor Items zur kognitiven Ebene, um mögliche Verärgerungen durch schwierige Wissensfragen nicht die affektive Bewertung beeinflussen zu lassen. Im Bereich zum kognitiven Wissen stehen die Items ohne Antwortvorgaben vor den Items mit Antwortvorgaben, um inhaltliche Anregungen, die die Schüler aus den Items mit Antwortvorgaben erhalten könnten, zu vermeiden. In den Nachtests wird nach den gleichen Grundsätzen verfahren. Nur das letzte Item (aus *Freie Äußerung*) sei hervorgehoben, nämlich „(...) zeichne (...), was Dir am Lindenhof am allerbesten gefallen hat.“ Dieses Item wird an das Ende des Tests gesetzt, da seine Bearbeitung viel Zeit in Anspruch nimmt und den Schülern Spaß macht (White & Gunstone 1992, S. 102). Ein unerwünschtes Zurückblättern und Nachbessern der Items durch die Probanden soll so möglichst vermieden werden.

Positionseffekte lassen sich trotz der oben beschriebenen Maßnahmen nicht ganz ausschließen (Bortz & Döring 1995, S. 514). Ein Nachteil von Positionseffekten liegt in der ungleichmäßigen Bearbeitung von Tests z.B. aufgrund von Ermüdung der Probanden. Dies dürfte bei Gymnasiasten der fünften Jahrgangsstufe, die routiniert sind, Tests von bis zu 60 min zu schreiben (GSO 2003, § 44/5), bei einem Fragebogen von maximal 20 min Dauer kein wichtiger Einflussfaktor sein. Innerhalb der einzelnen Testteile werden ferner die Items, soweit dies praktikabel und einleuchtend erscheint, variiert. Bei *KogMit25* wird z.B. die Reihenfolge der Items von Messzeitpunkt zu Messzeitpunkt vollständig verändert, um eine Erinnerung an die Position des jeweiligen Items im vorherigen Fragebogen zu erschweren.

Um absichtlich herbeigeführten Verfälschungen entgegenzuwirken, werden die Schüler aufgeklärt, dass die Testergebnisse keine Konsequenzen für sie haben, z.B. nicht in die Notengebung eingehen, so dass Simulation von Wissen keinen Vorteil bringt. „Dummstellen“ seitens der Schüler ist nicht zu erwarten, denn bei Schülern der fünften Jahrgangsstufe kann man i. d. R. von hinreichender Offenheit und positiver Hal-

tung gegenüber einem Museumsbesuch ausgehen, so dass Fragebögen nicht in erheblichem Maße absichtlich falsch oder unzureichend beantwortet werden. Diese hohe Kooperationsbereitschaft der Schüler (und der Lehrer) wird z.B. durch die sehr geringen Ausfallquoten dokumentiert. Von über 95% (366 von 385) der Schüler aus den 14 beteiligten Klassen liegen Vortest, Nachtest I und II in auswertbarer Form vor. Eine Anonymisierung der Fragebögen sollte der Informationskontrolle durch Selbstdarstellung, aber auch sozial erwünschter Beantwortungen entgegenwirken. Zur Vermeidung von Testverfälschungen durch Antworttendenzen, wie Akquieszenz (Ja-Sage-Tendenz) oder Nein-Sage-Tendenz, sind bei *KogMit25* etwa die Hälfte der Aussagen positiv (richtige Antwort lautet „stimmt“) und die Hälfte negativ (richtige Antwort lautet „stimmt nicht“) formuliert.

Verfälschungen durch zeitgebundene Störvariablen, gemeint sind externe Einflüsse, die während der Untersuchungszeit auftreten, und Reifungsprozesse der Probanden, werden als unerheblich eingeschätzt. Artefakte durch inhaltliche Beeinflussungen der Schüler innerhalb oder außerhalb des Unterrichts sind kaum zu erwarten, denn die betroffenen Inhalte sind während der Untersuchungszeit nicht Stoff des Lehrplans. Ferner werden die unterrichtenden Fachkollegen ausdrücklich gebeten, die in den Fragebögen angesprochenen Themen nicht zum Unterrichtsinhalt zu machen. Von besonderer Tagesaktualität war in der Untersuchungszeit keiner der im Museum behandelten Inhalte, so dass außergewöhnliche mediale bzw. öffentliche Präsenz eines Inhalts und damit eine unkalkulierbare Beeinflussung der Schüler nicht zu befürchten ist. Reifungsprozesse, die bei Probanden im Alter der beginnenden Pubertät eine gewichtige Rolle spielen können, sind bei einem Zeithorizont von nur wenigen Wochen eine zu vernachlässigende Größe.

5. Auswertung, statistische Behandlung und Darstellung der Daten

Die Kodierung wird wie folgt vorgenommen: Für die Fragebogenteile *AffMu*, *AffMuUr*, *AffMuPla* und „*Tieraffektivität*“ wird fehlende Kenntnis bzw. Nicht-Erinnern als 0, die übrigen Intensitäten von 1, geringe Intensität, bis 5, hohe Intensität, codiert, wobei bei den meisten Berechnungen die Kategorie 0 unberücksichtigt bleibt. Die Antworten zu den Items *KogO* werden von 0 bis ∞ (höchster von Schülern erreichter Wert ist 10) kodiert. Die Ausnahme stellt das Item zu „Platzbedarf“ dar. Hier erfolgt die Kodierung von 0 bis 4 mit 0=ganz falsch und 4=ganz richtig (präzisere Erläuterungen hierzu siehe später vor der Darstellung der Ergebnisse zu *KogO*). Items zu *KogMit* werden aus den zur Reliabilitätsanalyse angestellten Überlegungen kodiert als 1, richtige Antwort, und 0, übrige Kategorien; diese sind „falsch“ und „weiß nicht“. Die Items in *Freie Äußerung* erfordern Bewertungen oder inhaltliche Beschreibungen. Die sprachliche freie Bemerkung wird kodiert mit 0, keine, 1, positive, 2, negative, 3, positive und negative Äußerung, die zeichnerische Äußerung je nach Fragestellung mit 0, keine Zeichnung, und 1, angefertigte Zeichnung, oder 0, keine Zeichnung, und 1 bis ∞ , je nach Haupt- und Detailinhalten.

Die oben bereits erwähnten affektiven Items mit Antwortvorgabe bedienen sich v. a. fünfstufiger Likert-Skalen mit den Optionen „außerordentlich gut“, „ziemlich“, „mittelmäßig“, „kaum“, „gar nicht“ (vgl. Bortz & Döring 1995, S. 164), daneben der Möglichkeit einer Nicht-Antwort, um erzwungene Einschätzungen zu vermeiden. Diese Antwortmöglichkeiten beschreiben annähernd äquidistante Ausprägungen eines Merkmalskontinuums der Intensität. Dies wurde empirisch überprüft. (Bortz & Döring 1995, S. 164 ff., S. 203). Diese Äquidistanz macht es unproblematisch, die ordinalskalierten Daten in statistischen Tests wie intervallskalierte zu behandeln. Das entscheidende Kriterium, empirisch bedeutungsvolle Abstände zwischen den einzelnen Merkmalsausprägungen angeben zu können, ist gegeben (Bleymüller 2000, S. 3; Zöfel 2002, S. 14). Benninghaus (1991, S. 22 ff.), Allerbeck (1978, S. 199 bis 214) und Bortz & Döring (1995, S. 168 f.) befürworten sogar die Verwendung ordinaler Skalen als Intervallskalen. Für die verwendeten nichtparametrischen Tests reicht ordinalskaliertes Niveau der Daten aus (vgl. Zöfel 2002, S. 89). In Graphiken werden jedoch Mittelwerte verwendet, wofür die Daten intervallskaliert sein sollten.

Nachteile der dichotomen Struktur von *KogMit* sind die hohe Ratewahrscheinlichkeit und ein eventuelles Erzwingen von Antworten. Um dem letztgenannten potentiellen Schwachpunkt Rechnung zu tragen, wurde die Möglichkeit einer Nicht-Antwort („weiß nicht“) als dritte Alternative geboten. Das berührt die dichotome Entscheidung der antwortenden Probanden nicht, da eine Nicht-Antwort auf anderer Ebene anzusiedeln ist. Das Nicht-Erzwingen einer Antwort führt sicher z. T. dazu, dass Zustimmungstendenzen sich nicht in einer Entscheidung manifestieren. Die Ratewahrscheinlichkeit lässt sich rechnerisch bestimmen. Für die nach Bortz & Döring (1995, S. 196 ff.) korrigierte Ratewahrscheinlichkeit ergeben sich zwischen VT-NT I, VT-NT II, NT I und NT II die gleichen Verhältnisse wie bei den unkorrigierten Werten. Exemplarisch wurden einige Berechnungen zum Vergleich der korrigierten und unkorrigierten Werte durchgeführt. Es ergaben sich dieselben Zusammenhänge auf meist identischem Signifikanzniveau (Anhang 26). Das lässt diese Korrektur hinfällig erscheinen, und sie wurde darum zu Gunsten der Transparenz der Ausgangsdaten nicht verwendet.

Es werden Irrtumswahrscheinlichkeiten, statistische Signifikanzen, bestimmt, die Aussagen über die Wahrscheinlichkeit der Richtigkeit von Hypothesen zulassen (vgl. Wolf 2001, S. 97; Lamprecht 1999, S. 22 f.). Die statistische Auswertung der Daten erfolgt mit Hilfe des Computerprogramms SPSS Version 11.5. Folgende (willkürliche) Signifikanzgrenzen werden verwendet (vgl. Lamprecht 1999, S. 23 f.): Liegt $p > 0,05$, so gilt ein Zusammenhang als nicht signifikant „n.s.“, ab $p < 0,05$ als signifikant. $p < 0,01$ und $p < 0,001$ heißen hoch signifikant und höchst signifikant (verändert nach Zöfel 2002, S. 63).

Vor der Behandlung der zur Hypothesenüberprüfung verwendeten Prüfverfahren sei auf ein Problem hingewiesen, das einen Großteil der Auswertung betrifft. Die Überprüfung der Verteilung der Stichproben, Gesamtgruppe und Untergruppen, mittels Anpassungstest ergibt nicht in allen Fällen Normalverteilung (vgl. Anhang 27). Als Anpassungstest wird der Kolmogorov-Smirnov-Test verwendet. Man kann damit die Werte einer realen Stichprobe mit einer theoretischen Verteilung, hier der Normalverteilung, vergleichen (Zöfel 2002, S. 77 ff.). Als Skalenniveau genügen Nominaldaten. Im Gegensatz zum Chi-Quadrat-Test erlaubt der Kolmogorov-Smirnov-Test eine Verwendung der „individuellen Messwerte“ und kann auch bei kleinen Stichproben an-

gewendet werden, was bei der erwarteten Aufsplittung in Untergruppen einen Vorteil darstellt (vgl. Lamprecht 1999, S. 35).

In allen Fällen (normalverteilte und nicht-normalverteilte Stichproben) werden verteilungsfreie Prüfverfahren angewendet, damit innerhalb des gesamten Datensatzes bezüglich dieser Testebene Vergleichbarkeit gewährleistet ist. Die im Folgenden vorgestellten Tests erfordern keine Normalverteilung der Stichprobe. Diese Ebene der statistischen Tests ist zu trennen von der Ebene der Darstellung. Meist werden in den Abbildungen Mittelwerte dargestellt, wenngleich nichtparametrische Prüfverfahren Verwendung finden. Zu allen Ergebnissen sind im Tabellenanhang, dem letzten Anhangteil (Anhang 59) weitere statistische Größen wie Median, 1. & 3. Quartil, Streuungsmaße, etc. zusammengefasst.

„Der (...) [Friedmann]-Test dient zum Vergleich von mehr als zwei abhängigen Stichproben (...)“ (Zöfel 2002, S. 174). „Häufiger Anwendungsfall ist der, dass eine Messung zu verschiedenen Zeitpunkten vorgenommen wurde“ (Zöfel 2002, S. 174), z.B. ein Vergleich kognitiver Leistung zwischen Vortest, Nachtest I und II. Dabei klärt der Friedmann-Test nicht, „welche Zeitpunkte sich im einzelnen unterscheiden. Dies müsste man gegebenenfalls paarweise mit dem Wilcoxon-Test überprüfen (Zöfel 2002, S. 174). „Der (...) [Wilcoxon]-Test dient zum Vergleich zweier abhängiger Stichproben (...)“ (Zöfel 2002, S. 111), z.B. der kognitiven Leistung zwischen Vortest und Nachtest I.

Für statistische Vergleiche von mehr als zwei unabhängigen Variablen, z.B. beim Vergleich der drei Treatmentgruppen, ist der H-Test nach Kruskal und Wallis, im Folgenden als Kruskal-Wallis-Test bezeichnet, zu verwenden. „Der (...) [Kruskal-Wallis]-Test dient zum Vergleich von mehr als zwei Stichproben (...)“ (Zöfel 2002, S. 114). „Ergibt der [Kruskal-Wallis]-Test ein signifikantes Ergebnis ($p < 0,05$), so besagt es, dass die Nullhypothese ‚Alle Stichproben haben die gleiche zentrale Tendenz‘ verworfen werden muss“ (Zöfel 2002, S. 117). Für die Bestimmung von paarweisen Unterschieden führt man den U-Test nach Mann und Whitney aus, im Folgenden Mann-Whitney-U-Test genannt. „Der (...) [Mann-Whitney-U]-Test dient zum Vergleich von zwei Stichproben (...)“ (Zöfel 2002, S. 92).

Für diese vier Tests genügt jeweils Ordinalskalenniveau (Zöfel 2002, S 92).

Für Korrelationen zwischen zwei Variablen wird der Korrelationskoeffizient nach Spearman berechnet (Zöfel 2002, S. 126) mit der Einstufung der Koeffizienten nach Zöfel (2002, S. 120).

Bei nominalskalierten Daten, z.B. zwischen Zeichnungsinhalt und Treatmentgruppe, werden statistische Zusammenhänge mit Hilfe von Kreuztabellen mit Chi-Quadrat-Tests bestimmt. Bei sehr kleinen Häufigkeiten (<5) ist der exakte Test nach Fisher durchzuführen, im Folgenden Fisher-Test genannt (Zöfel 2002, S. 164).

Als Übereinstimmungskoeffizient wird Cohens Kappa (κ) bestimmt. „Im typischen Anwendungsfall werden Personen oder Objekte von zwei Beurteilern begutachtet. κ gibt den Grad der Übereinstimmung zwischen den Beurteilern an (...)“ (Zöfel 2002, S. 167). Nominales Skalenniveau ist ausreichend (Zöfel 2002, S. 89).

Für die Bestimmung der Reliabilität eines Tests ist „am gebräuchlichsten (...) der Alpha-Koeffizient von Cronbach“ (Bortz & Döring 1995, S. 184). Dichotome und polytome Items können mit dem Test behandelt werden. Der Test entspricht der mittleren Testhalbierungs-Reliabilität für alle Möglichkeiten einer Testhalbierung (Bortz & Döring 1995, S. 184).

Die Faktorenanalyse ist nach Zöfel (2002, S. 181 ff.) „ein Verfahren, das eine größere Anzahl von Variablen auf eine kleinere Anzahl hypothetischer Größen, Faktoren genannt, zurückführt. Diese Faktoren werden durch Variablengruppen gebildet (...)“. Lienert (1969, S. 490 ff.) beschreibt sie als ein Verfahren, das Aussagen darüber macht, ob Items eines Tests mehrere Dimensionen eines Konstrukts erfassen. Dafür werden Iteminterkorrelationen analysiert. In dieser Untersuchung wird die Faktorenanalyse zur Aufklärung der Dimensionen bzw. Faktoren der Items von Testteil *AffMu* eingesetzt. (Wie oben berichtet, lassen sich die sieben Items auf einen Faktor bzw. eine Dimension zurückführen.)

Während statistische Signifikanz ausdrückt, wie wahrscheinlich die Richtigkeit einer Hypothese ist, d.h. ob ein Effekt vorliegt, sagt die „praktische Signifikanz“ eines Effektstärkenmaßes aus, wie bedeutsam dieser Effekt ist (vgl. Wolf 2001, S. 97). Die standardisierte Effektstärke ω^2 nimmt Werte zwischen Null und eins an und „kann

(...) als prozentualer ‚Anteil der zugeschriebenen Varianz‘ interpretiert werden“ (Wolf 2001, S. 97). Der Wert $\omega^2=1$ beschreibt den theoretischen Fall, dass eine abhängige Variable gänzlich und fehlerfrei bestimmt wird, $\omega^2=0$ dagegen, dass die betreffende Variable gar keinen Einfluss hat. Für Metaanalysen lässt sich ω^2 in das Distanzmaß „d“ überführen (Wolf 2001, S. 98 f.), das z.B. in Häußler et al. (1998, S. 151 ff.) als Maß für die Effektstärke („Effektstärke ‚d‘“) verwendet wird.

Einfaktorielle Varianzanalysen werden in dieser Untersuchung ausschließlich zur Abschätzung der Effektgrößen von Treatment, Geschlecht und Bekanntheit des Lindenhofs verwendet. Der so ermittelte F-Wert erlaubt nach Wolf (2001, S. 99) eine Berechnung der Effektstärke ω^2 . Ebenso wird der t-Test ausschließlich zur Berechnung der Effektstärke des Museumsbesuchs eingesetzt. Mit Hilfe des t-Werts kann man nach Wolf (2001, S. 99) die Effektstärke ω^2 berechnen.

6. Untersuchungsdesign

Eine Woche vor dem Museumsbesuch schrieben die Schüler während des Schulunterrichts den Vortest. Direkt nach Durchführung der jeweiligen Treatments (S, SF, F) bearbeiteten sie noch im Museum den Nachtest I und nach 40 Tagen wiederum in der Schule den Nachtest II. Vergleichbare Studiendesigns weisen z.B. Reck (2000, S. 250 ff.), Leibold (1997, S. 84 f.), Grolnick & Ryan (1987) oder Randler (2002, S. 183) auf. Das Schreiben des Vortests in der Schule hat den organisatorischen Vorteil, während des Museumsbesuchs mehr Zeit zur Verfügung zu haben. Daneben sollte der zeitliche Abstand zwischen Vortest und Museumsbesuch die Wirkung des Vortests als nicht-intendierte, aber evtl. latent vorhandene Instruktion für den Museumsbesuch abschwächen. Der Nachtest I soll das durch den Museumsbesuch vermittelte Wissen widerspiegeln und wird darum sofort nach dem Besuch durchgeführt. Der erst nach 40 Tagen geschriebene Nachtest II soll dagegen langfristige Effekte der drei Treatments zeigen (vgl. Bogner 1998, 2004). Die Abbildung 39 veranschaulicht den zeitlichen Ablauf der Untersuchung und die eingesetzten Messinstrumente. Die Bearbeitungszeit für die Fragebögen betrug 15 Minuten für den Vortest (geringerer Umfang als Nachtests) und 20 Minuten für Nachtest I bzw. II. Wie aus den Vorstudien bekannt, reicht diese Zeitdauer für die Bearbeitung aus.

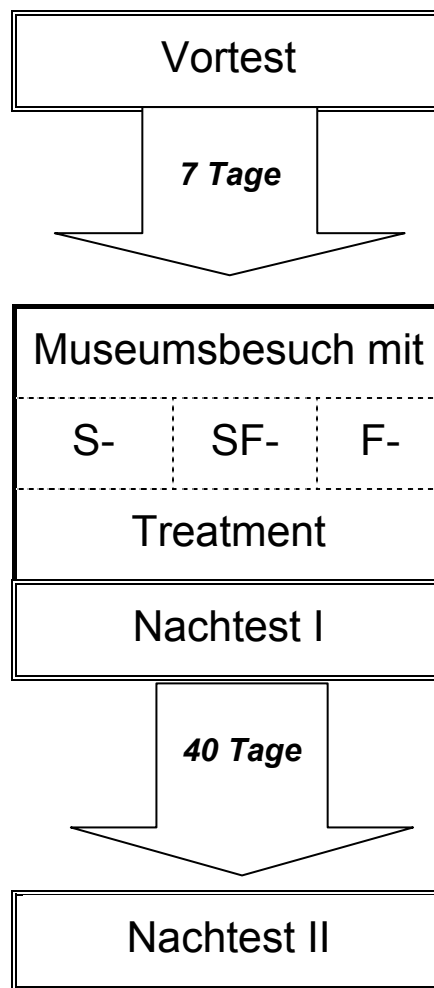


Abbildung 39: Untersuchungsdesign; Erläuterungen im Text.

7. Versuchsablauf

Für jeden Museumsbesuch und jede der 14 beteiligten Schulklassen galt der folgende Ablauf, wobei dieser Besuch eine verpflichtende Schulveranstaltung war, die am Vormittag an Stelle des regulären Unterrichts stattfand. Die Schüler wurden durch einen Brief informiert, der ihnen von ihrem Lehrer verlesen wurde. Sie wussten damit, dass sie einen Museumsbesuch durchführen werden und dass dieser Teil einer wissenschaftlichen Untersuchung ist, in der bestimmte Regeln einzuhalten sind (Anhang 28). Im Bestreben um Transparenz wird im Folgenden die Situation des Ausfüllens des ersten Fragebogens in der Schule möglichst genau geschildert. Nach dem Verlesen des Briefs wurde den Fünftklässlern der erste Fragebogen (Vortest) im Umfang von sieben Seiten mit 15 Minuten Bearbeitungszeit ausgeteilt. Das Ausfüllen

des Fragebogens wird somit durch den Brief angebahnt und mit „Positivem“ (Museumsbesuch statt Unterricht im Klassenraum) verknüpft. Dazu erfahren die Schüler, die zum Zeitpunkt der Untersuchung schon über ein halbes Jahr am Gymnasium sind und darum routiniert im Bearbeiten von Testaufgaben, vorab, dass dieser Test nicht in die Notengebung eingehen wird, sie darum frei von Prüfungsangst sein dürfen. Das Ausfüllen eines Fragebogens wird von Probanden als Kommunikation erlebt (Bortz & Döring 1995, S. 212). Diese Kommunikationssituation – begonnen durch das Verlesen des Briefs – wird fortgeführt mit sehr einfachen Erklärungen, wie z.B. zur Anonymisierung, und der Erhebung soziodemographischer Daten. Die Schüler werden persönlich angesprochen und geduzt, z.B. „Warst Du schon einmal in einem Naturkundemuseum?“ Einfache Erklärungen und Kommentare sollen den Schülern das Ausfüllen erleichtern, z.B. „Wenn Du eine Antwort nicht weißt, sollst Du nicht raten.“ oder „(...) So einfach geht's! Viel Spaß!“ So soll Distanz verringert und eine positive Grundhaltung geschaffen werden, um möglichst valide, d.h. v. a. von Selbstdarstellung und sozialer Erwünschtheit weitgehend freie Äußerungen der Probanden zu erhalten.

Am Tag des Museumsbesuchs trafen sich die Fünftklässler zunächst in der Schule, bevor sie mit einem Bus zum Lindenhof fuhren. Dort wurden sie vom Untersuchungsleiter in Empfang genommen und begrüßt. Es ist zu betonen, dass jegliche Intervention während des Museumsbesuches so weit wie möglich standardisiert wurde, so auch die Begrüßung der Schüler. Das heißt in diesem Fall, möglichst Wort für Wort ein vorgegebenes „Redekonzept“ einzuhalten. Nun erfolgte eine randomisierte Einteilung in die drei Treatmentgruppen. Die Schüler gingen sofort nach Zuteilung ihrer Gruppennummer, und damit Festlegung auf ein Treatment (1=S, 2=SF, 3=F), zu unterschiedlichen Sammelpunkten. Dort wurden sie von ihren Betreuerinnen erwartet. Es gab pro Treatmentgruppe mit etwa zehn Schülern eine Betreuerin. Diese erläuterte das weitere Vorgehen, reichte den Schülern Namensschilder, Stifte und Schreibblätter, teilte sie wiederum zufallsverteilt¹¹ in Kleingruppen von drei oder vier Probanden ein (eine Anzahl, die von Lou et al. (1996) zur Erreichung kognitiver Lerneffekte als die effektivste Gruppengröße ermittelt wurde) und gab ihnen zuletzt die entsprechenden Treatmentaufgaben (entsprechend ihrer Treatmentgruppe: S, SF, F).

¹¹ Zunächst wurden den Schülern randomisiert einseitig klebende farbige Kartons ausgeteilt, aus denen Namensschilder gefertigt wurden. Die für die Schüler zunächst bedeutungslose Farbe, zeigt die Zugehörigkeit zu einer Kleingruppe an.

Für den ganzen Museumsbesuch blieb die Betreuerin bei ihrer Gruppe, achtete auf die genaue Einhaltung der festgesetzten Bearbeitungszeiten und stellte sicher, dass die Schüler die gesamte zur Verfügung stehende Zeitspanne an der „richtigen“ Station verbrachten. Insgesamt bearbeiteten alle Schüler dieselben sechs Stationen im Naturkundemuseum und waren für eine halbe Stunde im Kindermuseum. Dies war durch einen Rotationsplan organisiert: Beispielsweise begann Treatmentgruppe S in Arbeitsphase I mit den Stationen 1 bis 3, bearbeitete nach kurzer Pause in Arbeitsphase II die Stationen 4 bis 6 und war nach der „großen Pause“ (20 min) in Arbeitsphase III für eine halbe Stunde im Kindermuseum. Wenige Minuten nach Arbeitsphase III und dem Einsammeln der Treatmentaufgaben folgte der Nachtest I (vgl. Tabelle 11). Dieser Nachtest fand in den Räumlichkeiten des Museums statt, jedoch getrennt von den bearbeiteten Stationen.

Tabelle 11: Exemplarische Darstellung einer Möglichkeit des Ablaufs des Museumsbesuchs. Tr.=Treatment, Aph=Arbeitsphase, St=Station, Ki=Kindermuseum, p=kleine Pause, P=große Pause; weitere Erläuterungen im Text.

Tr.	Aph I		Aph II		Aph III		
S	St 1-3	p	Ki	P	St 4-6	p	NT I
SF	St 4-6	p	St 1-3	P	Ki	p	NT I
F	Ki	p	St 4-6	P	St 1-3	p	NT I
Dauer	3 x 10 min	5 min	3 x 10 min	20 min	3 x 10 min	5 min	20 min

Innerhalb einer Arbeitsphase wechselten die Kleingruppen in analoger Weise (vgl. Anhang 29). Besonderer Wert wurde auf die Rotation der Betreuerinnen bezüglich der Treatments gelegt, so dass bei Vergleichen der Ergebnisse der Treatmentgruppen keine personenbezogenen Störungen zu berücksichtigen sind (vgl. Material und Methoden 1.). Anhang 4 belegt die Umsetzung des Betreuerinnenwechsels.

Im Kindermuseum wurden die Schüler jeweils für etwa eine halbe Stunde spielerisch beschäftigt. Hier wurden die Schüler der drei Treatmentgruppen nicht unterschiedlich behandelt. Folgende Aktivitäten fanden statt: Zunächst gab es ein Ratespiel, bei dem jedem Kind ein Schild mit einem Tiernamen auf den Rücken geheftet wurde. Die Kinder sollten „ihren“ Namen möglichst schnell und geschickt erfragen, wobei die Gruppe nur mit Ja und Nein antworten durfte. Danach wurde das erratene Tier im

Kindermuseum betrachtet. Der erfolgreiche Spieler durfte durch die Höhle krabbeln, wo hinter Glas mit schwacher Beleuchtung ein Dachs zu sehen war. Neben dem Dachs wurde bei Fuchs und Schwarzstorch Wert auf Vorkommen im Ratespiel und explizites Betrachten gelegt. Diese drei Tiere waren Teil des Fragebogens (*Kog-Mit25*). Nun erkletterten die Schüler das „Baumhaus“. Durch einen ausgehöhlten Stamm kann man den Giebel des Gebäudes erreichen, wo - wie im übrigen Kindermuseum – v. a. Stopfpräparate von Vögeln ausgestellt sind. Zuletzt wurde bis zum Ende dieser Arbeitsphase das „Kamera-Spiel“ durchgeführt. Zwei Schüler finden sich zusammen. Einer übernimmt den Part der Kamera, der andere den des Fotografen. Nun schließt der „Kamera-Schüler“ die Augen und lässt sich vom „Fotografen-Schüler“ lenken und auf ein Objekt ausrichten. Für einen Moment darf die „Kamera“ den Verschluss öffnen und den Film belichten bzw. der Schüler die Augen öffnen und sich das zu Sehende einprägen. Das Signal ist ein vorsichtiger Zug am Ohrläppchen. Nach einigen Spieldurchgängen, d.h. es wurden mehrere „Bilder gemacht“, zeigt die „Kamera“ dem „Fotografen“ das Aufgenommene. Die Schüler tauschen die Rollen und das Spiel beginnt erneut. Auf die genaue technische und inhaltliche Einhaltung des Ablaufs wurde geachtet. Andere Elemente des Kindermuseums wurden bewusst außer Acht gelassen, z.B. das Ecoboard (ein elektronisches Vogel-Ratespiel).

Nach dem Nachtest I durften die Schüler die Wartezeit bis zur Abfahrt im Museum verbringen. Fast alle Schüler erkletterten ein weiteres Mal das Baumhaus und verbrachten einen Großteil ihrer verbleibenden Zeit im Kindermuseum.

V. Ergebnisse

Im Folgenden werden zuerst die kognitiven Lernergebnisse des Museumsbesuchs unter den unterschiedlichen Treatmentbedingungen und unter Berücksichtigung methodischer Implikationen und der Bildung von Teilgruppen berichtet (vgl. Tabelle 12), anschließend in analoger Form die Ergebnisse zur affektiven Ebene (vgl. Tabelle 23).

1. Befunde zur kognitiven Aussageebene

Zur Klärung kognitiver Lernergebnisse werden Items mit und ohne Antwortvorgabe eingesetzt. Diese Übersicht soll als Orientierungshilfe dienen: Vergleiche Tabelle 12.

1.1. Resultate der Gesamtgruppe

1.1.1. Items mit Antwortvorgabe der kognitiven Aussageebene

Betrachtet werden zunächst die Ergebnisse zum Fragebogenteil *KogMit18*.

1.1.1.1. Kognitive Veränderungen in der Gesamtgruppe für die Items mit Antwortvorgabe

Für die Gesamtgruppe (N=366) interessiert die Lernleistung der Schüler im Nachtest I und II jeweils im Vergleich zum Vortest, demzufolge der kurzfristige Erwerb von Wissen und längerfristige Persistenz des Lernerfolgs. Weiter stellt sich die Frage, ob die im Nachtest I zu erwartenden Lernerfolge im Nachtest II auf dem gleichen Niveau bleiben, somit nach möglichem Wissensverlust bzw. nach dem Vergessen.

Es lassen sich höchst signifikante Unterschiede für die Ergebnisse zwischen den drei Testzeitpunkten nachweisen (Friedmann-Test; $p < 0,001$). Das höhere Ergebnis von Nachtest I und II ist jeweils im Vergleich zum Vortest statistisch bedeutsam, genauso wie der höhere Wert von Nachtest I verglichen mit Nachtest II (Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$; vgl. Abbildung 40). Die Gesamtgruppe der Schüler erwirbt Wissen (Vortest < Nachtest I), ein Teil ist persistent (Vortest < Nachtest II), ein Teil des erworbenen Wissens wird vergessen (Nachtest I > Nachtest II).

Tabelle 12: Übersicht zu den kognitiven Lernergebnissen der Gesamtgruppe und der Teilgruppen gebildet nach der Bekanntheit des Lindenhofs und nach dem Geschlecht.

Gesamtgruppe (1.1.)			
Items...			
...mit Antwortvorgabe (KogMit18) (1.1.1.)		...ohne Antwortvorgabe (KogO) (1.1.2.)	
<ul style="list-style-type: none">• Lernerfolg (1.1.1.1.)• Treatmentgruppenvergleich (1.1.1.2.)• Lernerfolg der Einzeltreatments (1.1.1.3.)• Itemtypenbezogene Effekte (1.1.1.4.)<ul style="list-style-type: none">• Zusammenfassung		<ul style="list-style-type: none">• (Methodische Vorbemerkung)<ul style="list-style-type: none">• Lernerfolg (1.1.2.1.)• Treatmentgruppenvergleich (1.1.2.2.)• Lernerfolg der Einzeltreatments (1.1.2.3.)<ul style="list-style-type: none">• Zusammenfassung	
Gegenüberstellung			
Teilgruppen (1.2.)			
Teilgruppen nach Bekanntheit des Lindenhofs (1.2.1.)			
Items...			
...mit Antwortvorgabe (KogMit18) (1.2.1.1.)		...ohne Antwortvorgabe (KogO) (1.2.1.2.)	
<ul style="list-style-type: none">• Vergleich zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern		<ul style="list-style-type: none">• Vergleich zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern	
Nicht-Kenner (1.2.1.1.1.)	Kenner (1.2.1.1.2.)	Nicht-Kenner (1.2.1.2.1.)	Kenner (1.2.1.2.2.)
	Vorerfahrungen der Kenner		
Lernerfolg	Lernerfolg	Lernerfolg	Lernerfolg
Treatmentgruppenvergleich	Treatmentgruppenvergleich	Treatmentgruppenvergleich	Treatmentgruppenvergleich
Lernerfolg der Einzeltreatments	Lernerfolg der Einzeltreatments	Lernerfolg der Einzeltreatments	Lernerfolg der Einzeltreatments
Itemtypenbezogene Effekte	Itemtypenbezogene Effekte		
Zusammenfassung	Zusammenfassung	Zusammenfassung	
Gegenüberstellung und Zusammenfassung			
Teilgruppen nach Geschlecht der Probanden (1.2.2.)			
Items...			
...mit Antwortvorgabe (KogMit18) (1.2.2.1.)		...ohne Antwortvorgabe (KogO) (1.2.2.2.)	
<ul style="list-style-type: none">• Lernerfolg: Gesamtgruppe, Nicht-Kenner und Kenner		<ul style="list-style-type: none">• Lernerfolg: Gesamtgruppe, Nicht-Kenner und Kenner	
Zusammenfassung			

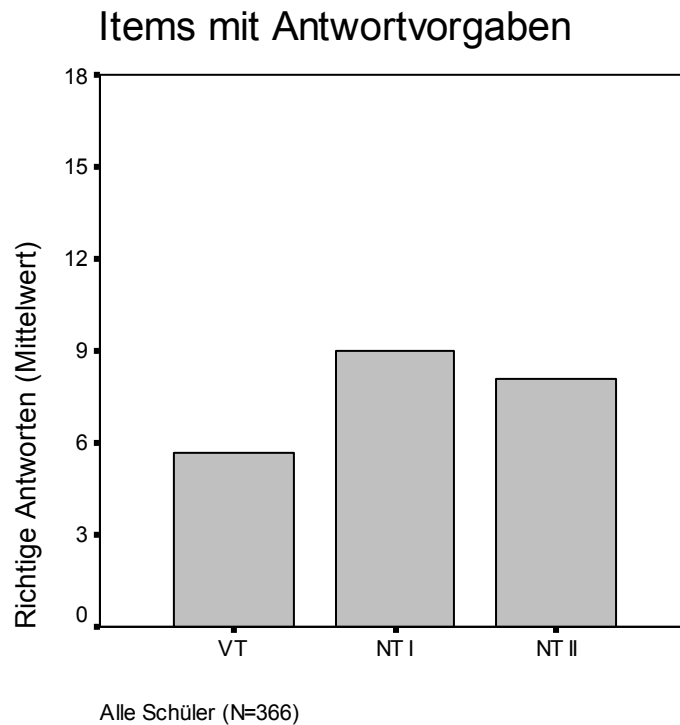


Abbildung 40: Lernerfolg aller Schüler bei *KogMit18*. Erläuterungen im Text.

Diese Sprachregelung bezüglich der „Wissensveränderung“ der Schüler soll künftig beibehalten werden. Nur signifikante Unterschiede werden berücksichtigt (mindestens $p < 0,05$).

Erwerb:	Vortest < Nachtest I
Persistenz:	Vortest < Nachtest II
Vergessen:	Nachtest I > Nachtest II

1.1.1.2. Vergleich zwischen den Treatmentgruppen

Ein Hauptziel der Untersuchung besteht in der Bestimmung der unterschiedlichen Lerneffektivität der drei Treatmentgruppen (S, SF, F). Welche Art des Treatments zeitigt besseren Lernfortschritt: das selbstgesteuerte Arbeiten (S), das fremdgesteuerte (F) oder die Mischform (SF)?

Vorab ist die fast völlige Gleichheit der drei Ausgangswerte im Vortest zu bemerken (Kruskal-Wallis-Test; $p = 0,796$; vgl. Abbildung 41). Dies ist als Voraussetzung für einen Vergleich der Treatmentgruppen sehr günstig. Für beide Messungen Nachtest I

und II lassen sich statistisch bedeutsame Differenzen nachweisen (Kruskal-Wallis-Test; Nachtest I: $p < 0,001$, Nachtest II: $p = 0,002$). Überprüft man diese, so ergeben sich im Nachtest I höchst signifikante Unterschiede zwischen S- und SF-Schülern zu Gunsten von Treatment SF, hoch signifikante Unterschiede zwischen S- und F-Schülern zu Gunsten von Treatment F und signifikante Unterschiede zwischen SF- und F-Schülern zu Gunsten von Probanden des Treatments SF (Mann-Whitney-U-Test; $p(S < SF) < 0,001$, $p(S < F) = 0,005$, $p(SF > F) = 0,021$). Im Nachtest II bleiben nur die Signifikanzen zwischen den Paarungen S/SF und SF/F bestehen (Mann-Whitney-U-Test; $p(S < SF) = 0,001$, $p(S/F) = 0,406$, $p(SF > F) = 0,008$).

Die Lernwirksamkeit der Treatments ist anscheinend unterschiedlich. Die Mischform SF schneidet besonders gut ab.

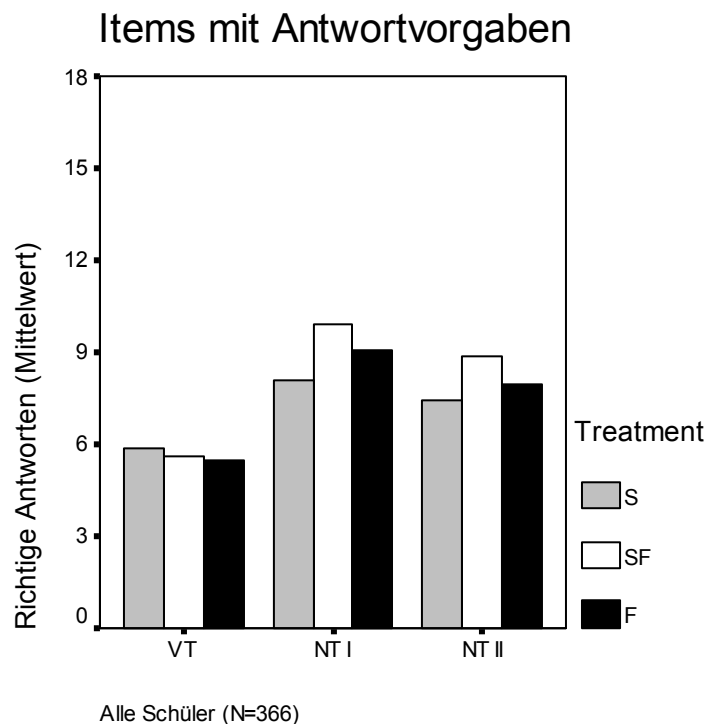


Abbildung 41: Vergleich zwischen den Treatmentgruppen bei der Gesamtgruppe bei *KogMit18*. Erläuterungen im Text.

1.1.1.3. Kognitive Veränderungen der Treatmentgruppen

Diese Unterschiede erfordern eine genauere Betrachtung auf der Ebene der einzelnen Treatmentgruppen. Lassen sich für die einzelnen Treatmentgruppen, so wie gerade für die Gesamtgruppe gezeigt, Erwerb, Persistenz und Vergessen von Wissen nachweisen?

F-Schüler: Wie Abbildung 41 zeigt, kann die obige Frage für die Teilgruppe, die das F-Treatment bearbeitete, positiv beantwortet werden. Die Schüler gewinnen Wissen hinzu, ein Teil bleibt ihnen erhalten, ein Teil wird eingebüßt (Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(VT < NTI) < 0,001$, $p(VT < NTII) < 0,001$, $p(NTI > NTII) = 0,019$; $N = 122$).

SF- und S-Schüler: Entsprechendes gilt für das Treatment SF und S (beide: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$; $N(\text{SF-Schüler}) = 120$; $N(\text{F-Schüler}) = 124$; Abbildung 41).

Für alle drei Treatmentgruppen gilt: Die Schüler zeigen Erwerb, Persistenz und Vergessen von Wissen.

1.1.1.4. Kognitive Veränderungen bezogen auf die Itemtypen

Die 18 Items mit Antwortvorgabe aus KogMit18 beziehen sich inhaltlich auf die sechs von den Schülern bearbeiteten Schaukästen bzw. Dioramen. Trotz dieser Begrenzung und weitgehender inhaltlicher Übereinstimmung zwischen „Museum und Fragebogen“ wird auf noch feinerer Ebene der Inhalt des Fragebogens kontrolliert. Für jedes Schaufenster und für jedes der drei Treatments wird ein Item bezogen auf die jeweiligen Schwerpunkte des Treatments konstruiert, so dass pro Treatment sechs, insgesamt 18 Items mit Antwortvorgaben Teil des Fragebogens sind. Es gibt F-Items, die so konstruiert sind, dass sie wahrscheinlich besonders gut von Schülern, die das F-Treatment durchlaufen haben, beantwortet werden können; analog gibt es SF- und S-Items. Damit werden soweit wie möglich Bevorzugungen einer der Treatmentgruppen ausgeschlossen (Näheres vgl. Material & Methoden Kapitel 4.2.1.).

Diese Voraussetzungen erlauben eine noch detailliertere Analyse der Ergebnisse. Man kann nämlich nun die jeweils sechs F-, SF- und S-Items gesondert für die Treatmentgruppen der F-, SF- und S-Schüler überprüfen. Besonders wichtig ist an dieser Stelle, zwischen den Items, die für eine Treatmentgruppe konstruiert wurden, und der Treatmentgruppe selbst zu unterscheiden.

Zunächst zu den F-Items (vgl. Abbildung 42): Auffällig ist das besonders gute Abschneiden der F-Schüler im Nachtest I. Die Differenzen sind statistisch bedeutsam. Im Nachtest II sind F- und SF-Schüler gleichauf. Nur die S-Schüler sind schlechter. (Vortest: Kruskal-Wallis-Test; $p=0,934$; Nachtest I: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S<SF)=0,040$, $p(S<F)<0,001$, $p(SF<F)<0,001$; Nachtest II: Kruskal-Wallis-Test; $p=0,008$; Mann-Whitney-U-Test; $p(S<SF)=0,003$, $p(S<F)=0,020$, $p(SF/F)=0,527$).

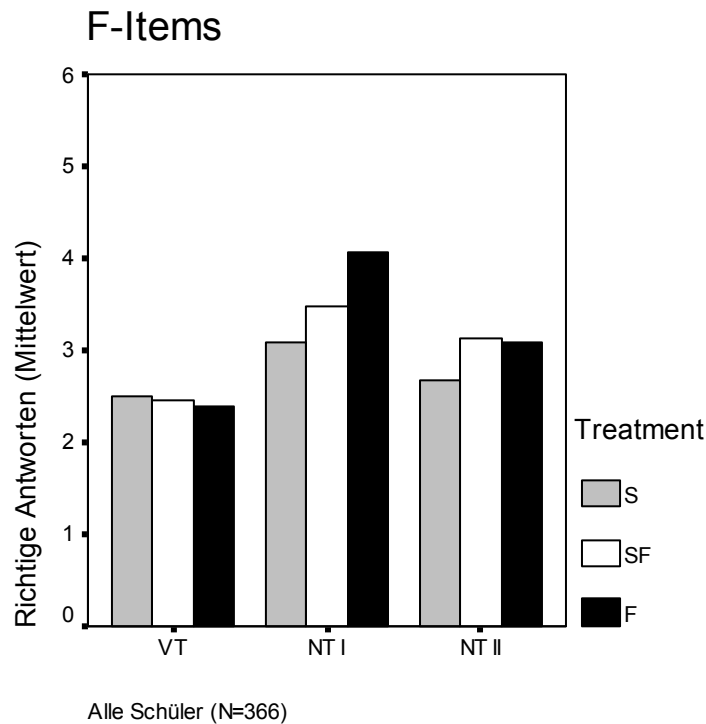


Abbildung 42: Lernerfolg aller Schüler bezüglich der F-Items bei *KogMit18*. Erläuterungen im Text.

SF-Items (vgl. Abbildung 43): Die Gruppe der SF-Schüler schneidet im Nachtest I und II signifikant besser ab als die übrigen Gruppen (Vortest: Kruskal-Wallis-Test; $p=0,837$; Nachtest I: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S<SF)<0,001$, $p(S/F)=0,863$, $p(SF>F)<0,001$; Nachtest II: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S<SF)<0,001$, $p(S/F)=0,081$, $p(SF>F)=0,001$).

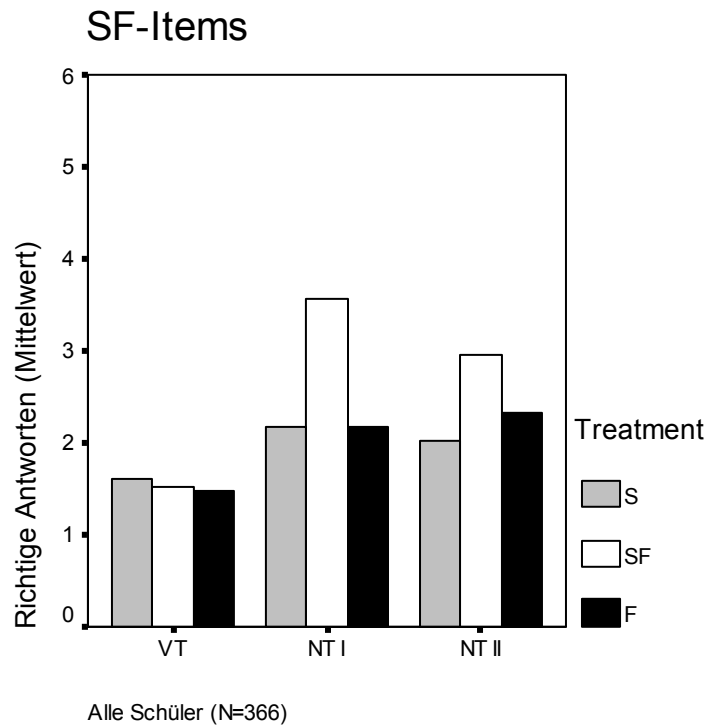


Abbildung 43: Lernerfolg aller Schüler bezüglich der SF-Items bei *KogMit18*. Erläuterungen im Text.

S-Items (vgl. Abbildung 44): Zwar erreichen die S-Schüler bei diesen Items verglichen mit F- und SF-Items die relativ besten Ergebnisse (vgl. Abbildungen 42 und 43), dennoch sind sie in Nachtest I und II nur in etwa gleich gut wie die F- und SF-Schüler (vgl. Abbildung 44). Keiner der Unterschiede bei S-Items zwischen den Treatmentgruppen ist statistisch abzusichern (Vortest: Kruskal-Wallis-Test; $p=0,629$, Nachtest I: Kruskal-Wallis-Test; $p=0,895$; Nachtest II: Kruskal-Wallis-Test; $p=0,190$).

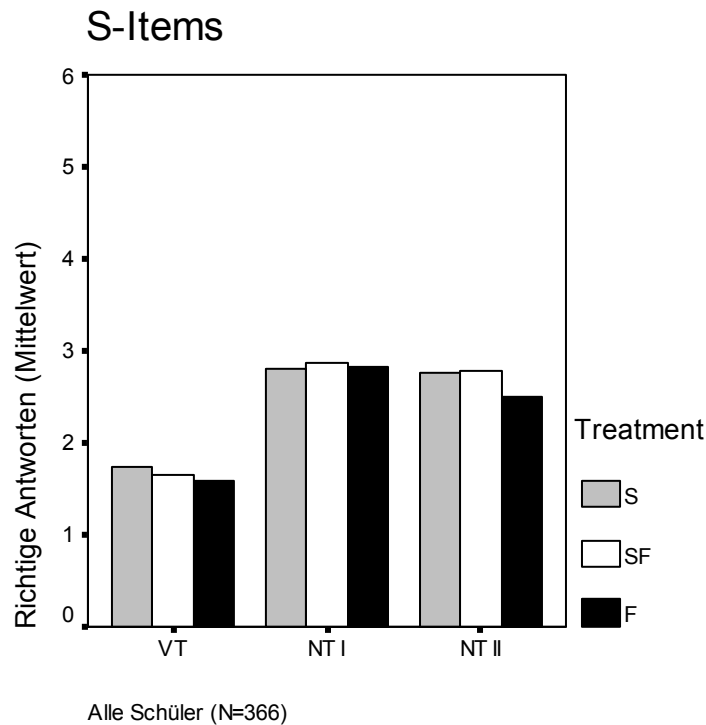


Abbildung 44: Lernerfolg aller Schüler bezüglich der S-Items bei *KogMit18*. Erläuterungen im Text.

Die Itemkonstruktion hat demnach Einfluss auf das Abschneiden der Treatmentgruppen. Die Werte in dem Wissenstest der jeweils in der Itemkonstruktion besonders beachteten Gruppe sind mindestens genauso hoch, meist signifikant höher als die der übrigen. Von dieser treatmentbezogenen Fragebogenkonstruktion profitieren die SF-Schüler am meisten, die S-Schüler am wenigsten.

Zusammenfassung zur kognitiven Aussageebene der Gesamtgruppe bei Items mit Antwortvorgabe:

- Die Schüler erwerben Wissen, behalten einen Teil davon und vergessen etwas.
- Das gilt für alle Treatmentgruppen: F-, SF- und S-Schüler.
- SF-Schüler haben die besten Ergebnisse.
- Die Itemkonstruktion beeinflusst die Ergebnisse: SF-Schüler haben bei SF-Items relativ bessere Ergebnisse als die übrigen Treatmentgruppen bei „ihren“ jeweiligen Items.

1.1.2. Items ohne Antwortvorgabe der kognitiven Aussageebene

Die bisherigen Ergebnisse beziehen sich auf Items mit, die folgenden auf Items ohne Antwortvorgabe (Fragebogenteil KogO). Zum Erhalt eines Gesamtbildes kognitiven Lernerfolgs ist es hilfreich, sich möglichst unterschiedlicher Messinstrumente zu bedienen.

Exkurs: Vor der Vorstellung dieser Ergebnisse ist es erforderlich, eine methodische Frage bewusst zu machen: Die Items ohne Antwortvorgaben setzen sich zusammen aus Items, die beschreibende (Urwald, Wacholderheide, Sandgrube und Kulturfolger), aufzählende (Archaeopteryx) und erläuternde (Platzbedarf) Äußerungen der Probanden bedingen und entsprechend unterschiedlich ausgewertet werden, nämlich Bewertungspunkte von 0-∞ für Aufzählungen und Beschreibungen und für Erklärungen von 0-4. Aus pragmatischen Gründen ist eine Zusammenfassung dieser Werte unterschiedlicher Dimension in Form von Summenscores anzustreben. Die Dimensionen stellen ähnliche Größen mit vergleichbaren Funktionen dar. Durch sie sollen Wissenszustände bewertet werden. Die Werte liegen in vergleichbarer Größenordnung, wobei die größten Unterschiede zwischen den formal ähnlichsten Typen bestehen (Mittelwerte / Nachtest I: Wacholderheide: 1,25; Urwald: 2,30; vgl. Tabelle 13). Die Maxima sind entweder vorgegeben (bei Platzbedarf), oder sie werden durch das jeweils beste von Schülern erbrachte Ergebnis bestimmt (alle übrigen).

Tabelle 13: Ergebnisse der Items ohne Antwortvorgaben. Dargestellt sind die Mittelwerte aller Schüler zu den unterschiedlichen Stationen.

	Vortest	Nachtest I	Nachtest II	Maxima
Urwald	1,73	2,30	2,08	10
Wacholderheide	0,22	1,25	0,90	9
Sandgrube	0,72	1,70	1,65	7
Kulturfolger	1,72	2,01	1,98	8
Archaeopteryx	0,55	1,57	1,04	7
Platzbedarf	0,29	1,81	1,69	4

Folgende Möglichkeiten bieten sich an, aus diesen Daten Summenscores zu bilden:

- a. Addition ohne Veränderung der Originaldaten, d.h. Akzeptieren der Unschärfen, die sich aus der Addition der letztlich willkürlich gewählten Bewertungen (insbesondere bezüglich der Station Platzbedarf) ergeben.*
- b. Orientierung an identischen Punktemaxima für alle Items unabhängig von der Art der Aufgabe und nachfolgender Addition.*
- c. Addition nach durchgeführter Normierung auf Richtwerte zu bestimmten Messzeitpunkten oder nach Normierung orientiert an Maxima.*

Obwohl im Ergebnisteil keine Methoden zu beschreiben sind, sei in diesem Fall ein methodischer Exkurs erlaubt, da die Frage nach einer Bildung von Summenscores nur für dieses Messinstrument relevant ist und die Ergebnisse deutlich beeinflussen kann. Die unterschiedlichen Verfahren werden kurz umrissen.

Für die Wahl der ersten Möglichkeit spricht die relativ größte Transparenz des Verfahrens. Die methodische Schwäche ist leicht erkennbar und einschätzbar. Ein richtiger „Inhalt“, ganz gleich, ob beschreibend oder aufzählend, wird als ein Punkt bewertet. Die Skala ist nicht nach oben beschränkt. Die Antworten zu Platzbedarf werden von 0 („völlig falsch“) bis 4 („ganz richtig“) eingestuft, dann wie die übrigen Punktwertungen behandelt. Der Mittelwert im Nachtest I von Platzbedarf beträgt 1,81, liegt damit im Mittelfeld der Mittelwerte der übrigen Items (vgl. Tabelle 13).

Das zweite Verfahren im Vorhinein für jedes Item ein maximal zu erreichendes Niveau durch identische Punktemaxima festzulegen, ist für die verwendeten Items z. T. problematisch. Beispielsweise für die Frage „Wie sieht für Dich der Urwald im Fichtelgebirge aus?“ kann es kein vorher bestimmbares optimales Ergebnis, keine Musterlösung geben. Jede Festsetzung von „schon ausreichend umfangreich beschrieben“ bzw. „nicht ausreichend beschrieben“ kann nur willkürlich sein, und ist den Schülern zum Zeitpunkt der Bearbeitung nicht bekannt und sollte den Schülern nicht bekannt sein, um Verfälschungen durch diesen externen Faktor zu vermeiden. Dasselbe gilt für jegliche gestufte Bewertung. Eine nachvollziehbare inhaltliche Begründung allein aufgrund unterschiedlicher Anzahl richtig genannter Fakten ist nicht zu sehen.

Für das dritte Verfahren bieten sich unterschiedliche Möglichkeiten an: Man kann alle Daten auf den Durchschnitt eines Messzeitpunktes normieren. Wählt man den Vortest, so betont man die vorherige Kenntnis bzw. Unkenntnis bestimmter Inhalte sehr stark. Wacholderheide, vor dem Treatment kaum bekannt, ergibt im Vortest im Mittel 0,22 sinnvolle und richtige Nennungen, Urwald dagegen 1,73. Damit würden die Daten von Wacholderheide bei einer Normierung auf den Vortest relativ zu Urwald fast das achtfache Gewicht bekommen. Jeder bei dem Item zur Wacholderheide auftretende Effekt würde damit das Gesamtergebnis um ein Vielfaches mehr beeinflussen als das Item zum Urwald. Das ist weder durch die zeitlichen Anteile während des Museumsbesuchs, noch durch jegliche weitere Umstände oder Ergebnisse begründbar und erscheint darum nicht gerechtfertigt. Eine Normierung auf andere Zeitpunkte birgt dasselbe Grundproblem.

Mit dem Argument, in dieser Stichprobe sei die „Musterlösung“ durch das beste erzielte Ergebnis bestimmt, kann man alle Daten auf die jeweils bei den unterschiedlichen Items erreichten Maxima beziehen und auf das insgesamt beste Ergebnis normieren. Der höchste erreichte Wert liegt bei zehn richtigen Inhalten (Urwald). Damit ist dies der Referenzwert. Die Daten der übrigen Items werden entsprechend der Maxima bezüglich dieses Referenzwertes bearbeitet. Bei dem Maximum sieben (Sandgrube) wird beispielsweise jeder Messwert der „Rohdaten“ dieses Items mit dem Faktor $10/7$ multipliziert. Die fünf Stufen bei „Platzbedarf“ (0 bis 4) werden ebenfalls entsprechend umgerechnet, so dass folgende Punktestufung resultiert: $0 \times 10/4 - 1 \times 10/4 - 2 \times 10/4 - \text{etc.} \rightarrow 0 - 2,5 - 5 - 7,5 - 10$. Diese Normierung auf Ergebnismaxima ergibt ein Übergewicht des Items Platzbedarf, zu erklären durch eine größere Datenhäufung in der Nähe des Maximums bei diesem Item im Vergleich mit den übrigen. Zudem wird der Umrechnungsfaktor allein durch den jeweiligen maximalen Wert eines Schülers von 366 Schülern festgesetzt. Das Verfahren ist somit sehr stark vom Zufall der Auswahl der Stichprobe beeinflusst.

Aus den unterschiedlichen vorgestellten und diskutierten Möglichkeiten wird das erste Verfahren ausgewählt, d.h. die Bildung der Summenscores ohne jede zusätzliche Veränderung bzw. Verfremdung der Daten. Obwohl dieses Verfahren Schwächen hat, herrscht mit seiner Verwendung die größte Transparenz. Keines der übrigen Verfahren kann mehr überzeugen.

Als Kontrolle wird ein weiteres Verfahren, die oben beschriebene Normierung anhand der Maxima, umgesetzt und die entsprechenden Werte berechnet. Ein Vergleich bezüglich der wichtigsten Punkte zeigt, dass die Aussagen nicht divergieren (vgl. Anhang 30).

1.1.2.1. Kognitive Veränderungen in der Gesamtgruppe für Items ohne Antwortvorgabe

Für die Gesamtheit aller Schüler gilt analog zu den Ergebnissen der Items mit Antwortvorgaben, die Schüler erwerben Wissen, behalten einen Teil und vergessen einen Teil wieder (Friedmann-Test, $p < 0,001$; Wilcoxon-Test, alle zu vergleichenden Paare $p < 0,001$; vgl. Abbildung 45).

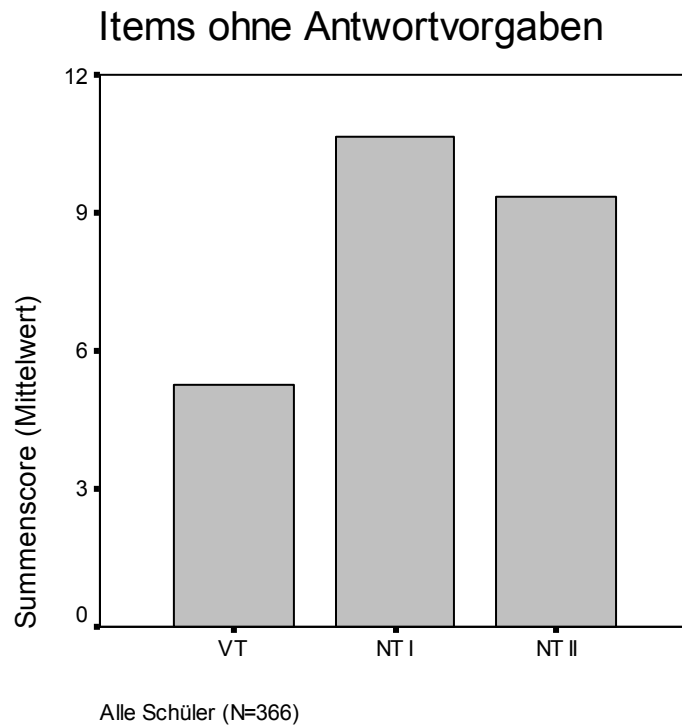


Abbildung 45: Lernerfolg aller Schüler bei *KogO*. Erläuterungen im Text.

1.1.2.2. Vergleich zwischen den Treatmentgruppen

Im Vortest und Nachtest II zeigen die Ergebnisse der drei Treatments keine Differenzen (Kruskal-Wallis-Test; $p(\text{VT})=0,940$; $p(\text{NT II})=0,259$; vgl. Abbildung 46). Statistische Unterschiede zwischen den Treatmentgruppen zeigen sich nur im Nachtest I (Kruskal-Wallis-Test; $p=0,001$). Ganz im Gegensatz zu den Ergebnissen von *KogMit18* schneiden hier die S-Schüler besser ab als die SF- und die F-Schüler (Mann-Whitney-U-Test; $p(\text{S}>\text{SF})=0,002$, $p(\text{S}>\text{F})=0,001$). Die Ergebnisse von Treatment SF und F sind praktisch identisch (Mann-Whitney-U-Test; $p(\text{SF}/\text{F})=0,817$).

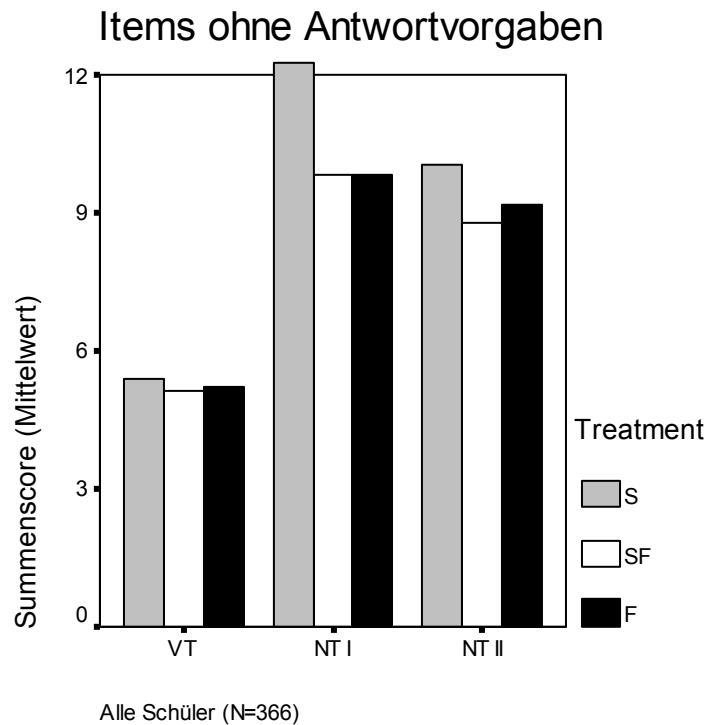


Abbildung 46: Vergleich zwischen den Treatmentgruppen bei der Gesamtgruppe bei *KogO*. Erläuterungen im Text.

1.1.2.3. Kognitive Veränderungen der Treatmentgruppen

Die Analyse der Einzeltreatments ergibt Überraschendes (vgl. Abbildung 46): In allen Treatmentgruppen unterscheiden sich die Messwerte (Friedmann-Test; in allen Gruppen $p=0,001$). Während sich bei S- und SF-Schülern Erwerb, Persistenz und Vergessen von Wissen zeigen lassen (beide Teilgruppen: Friedmann-Test; $p<0,001$; S-Schüler: Wilcoxon-Test: $p(VT<NT\ I)<0,001$, $p(VT<NT\ II)<0,001$, $p(NT\ I>NT\ II)<0,001$. SF-Schüler: Wilcoxon-Test: $p(VT<NT\ I)<0,001$, $p(VT<NT\ II)<0,001$, $p(NT\ I>NT\ II)=0,002$), kann man bei F-Schülern Erwerb und Persistenz, nicht aber Vergessen finden (Friedmann-Test; $p<0,001$; Wilcoxon-Test: $p(VT<NT\ I)<0,001$, $p(VT<NT\ II)<0,001$, $p(NT\ I-NT\ II)=0,132$; $N(F\text{-Schüler})=122$).

Treatmentbezogene Itemkonstruktion spielt bei *KogO* keine Rolle.

Zusammenfassung zur kognitiven Aussageebene der Gesamtgruppe bei Items ohne Antwortvorgabe:

- Die Schüler erwerben Wissen, behalten einen Teil davon und vergessen etwas.
- Im Gegensatz zu S- und SF-Schülern vergessen F-Schüler nicht.
- S-Schüler haben die besten Ergebnisse.

1.1.3. Gegenüberstellung der Hauptergebnisse der Items mit Antwortvorgabe (*KogMit18*) und der Items ohne Antwortvorgabe (*KogO*)

Die Grundaussage bleibt bei beiden Testverfahren dieselbe. Der Museumsbesuch führt zu Wissenserwerb, der selbst nach 40 Tagen noch nachweisbar ist. Je nach Testverfahren schneiden unterschiedliche Treatments besonders gut ab: Die Mischform zwischen Selbstbestimmung und Fremdbestimmung (SF-Treatment) führt im Hauptteil des Fragebogens mit Multiple-Choice-Items (*KogMit18*) zu den besseren Resultaten, dagegen haben selbstbestimmte Lerner (S-Treatment) bei Items, die selbstformulierte Antworten einfordern (*KogO*), die besseren Ergebnisse (Details siehe Übersichtsdarstellung Tabelle 14).

1.2. Resultate von Teilgruppen

Aus einer Vielzahl von Teilgruppen, die begründet differenziert ausgewertet werden können, haben sich in dieser Untersuchung folgende als besonders bedeutsam herausgestellt: Bekanntheit des Lindenhofs, Geschlecht der Probanden.

1.2.1. Bekanntheit des Lindenhofs

Mitentscheidend für das Ergebnis der Schüler in den Wissenstests zu den Museumsinhalten könnte die Kenntnis des Museums sein. Darum wird nun das obige Ergebnis zum Lernerfolg der Gesamtgruppe nach der Bekanntheit des Lindenhofs bei den Schülern vor der Durchführung dieser Untersuchung differenziert.

Tabelle 14: Übersicht über die Gegenüberstellung der Items mit Antwortvorgabe und der Items ohne Antwortvorgabe mit bilanzierender Wertung. „+“=“trifft zu“, „-“=“trifft nicht zu“, „G“=“gleich bzw. in der Kernaussage kein Unterschied“, „U“=“ungleich bzw. qualitativer Unterschied in der Aussage“. Weitere Erläuterungen im Text. N=366 oder wie angegeben.

Lernerfolg	Items mit Antwortvorgabe	Items ohne Antwortvorgabe	Bilanz
	Gesamtgruppe, N=366	Gesamtgruppe, N=366	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	+	G
	S, N=124	S, N=124	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	+	G
	SF, N=120	SF, N=120	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	+	G
	F, N=122	F, N=122	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	-	U
Treatment- unterschiede	Items mit Antwortvorgabe	Items ohne Antwortvorgabe	Bilanz
	Gesamtgruppe, N=366	Gesamtgruppe, N=366	
VT	Keine Unterschiede	Keine Unterschiede	G
NT I	SF>F>S	S>F=SF	U
NT II	SF>F=S	Keine Unterschiede	U

1.2.1.1. Items mit Antwortvorgabe (*KogMit18*)

Schüler, die schon mindestens einmal den Lindenhof besucht haben und der ihnen von unterschiedlichen Aktivitäten (u. a. Museumsbesuchen) bekannt sein kann, haben wahrscheinlich besseres Vorwissen als die übrigen und schneiden deshalb in den Tests besser ab. Dies ist zu testen.

Zu allen Messzeitpunkten lassen sich die vermuteten Unterschiede zu Gunsten der Lindenhof-Kenner statistisch absichern (Mann-Whitney-U-Test; VT: $p=0,032$, NT I: $p=0,018$, NT II: $p<0,001$; vgl. Abbildung 47): Die vorherige Bekanntheit der Einrichtung erbringt höhere Resultate im Wissenstest sowohl in Vor- als auch in Nachtests.

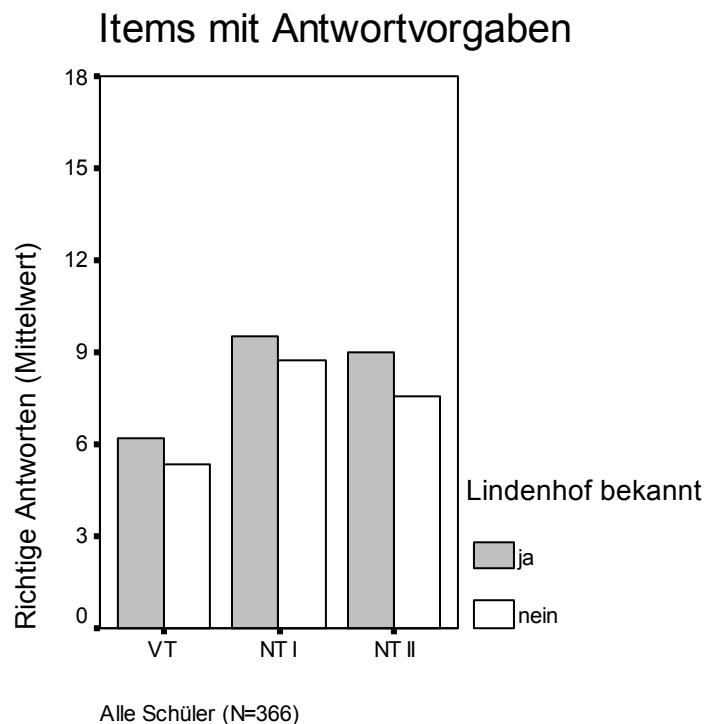


Abbildung 47: Vergleich des Lernerfolgs der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner bei *KogMit18*. Erläuterungen im Text.

Für eine detaillierte Analyse der entscheidenden Einflussfaktoren des Lernerfolgs müssen die Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner getrennt betrachtet werden.

1.2.1.1.1. Ergebnisse der Lindenhof-Nicht-Kenner für Items mit Antwortvorgabe (*KogMit18*)

Im Folgenden werden die Einzelergebnisse der Gruppe der 241 Schüler dargestellt, die den Lindenhof vor dieser Untersuchung **nicht** gekannt haben.

Für den Lernerfolg dieser Teilgruppe zeigt sich ein ähnliches Bild wie bei der Gesamtgruppe (vgl. Abbildung 48). Die Schüler gewinnen Wissen hinzu, mindestens ein Teil davon ist persistent und ein Teil geht wieder verloren (Friedmann-Test; $p < 0,001$; Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$).

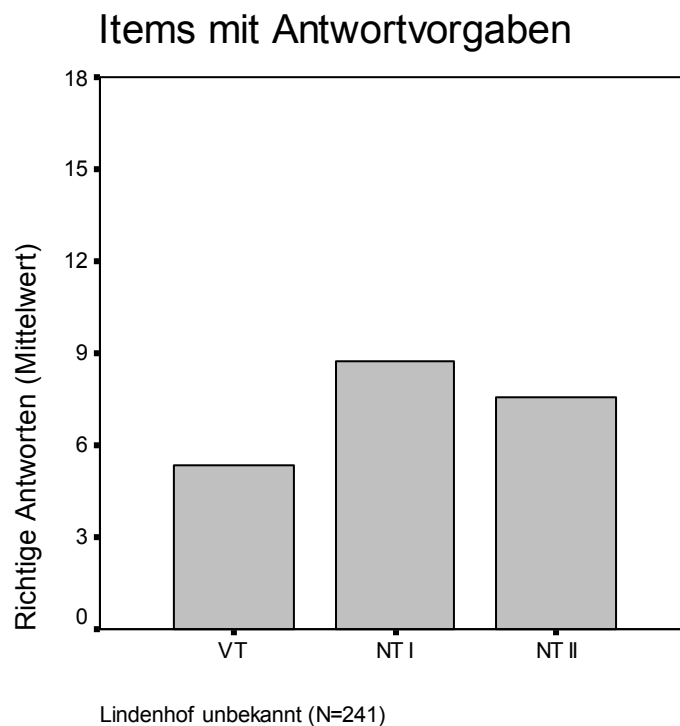


Abbildung 48: Lernerfolg der Lindenhof-Nicht-Kenner bei *KogMit18*. Erläuterungen im Text.

Der Treatmentvergleich (vgl. Abbildung 49) erbringt keine Unterschiede im Vortest (fast identische Resultate der F-, SF- und S-Schüler: Kruskal-Wallis-Test; $p = 0,998$) und mindestens hoch signifikante Unterschiede in Nachtest I und II (Kruskal-Wallis-Test; NT I: $p < 0,001$, NT II: $p = 0,005$). SF-Schüler schneiden am besten ab, S-Schüler am schlechtesten (Mann-Whitney-U-Test; NT I: $p(S < SF) < 0,001$, $p(S < F) = 0,003$, $p(SF > F) = 0,031$; NT II: $p(S < SF) = 0,003$, $p(S < F) = 0,522$, $p(SF > F) = 0,010$).

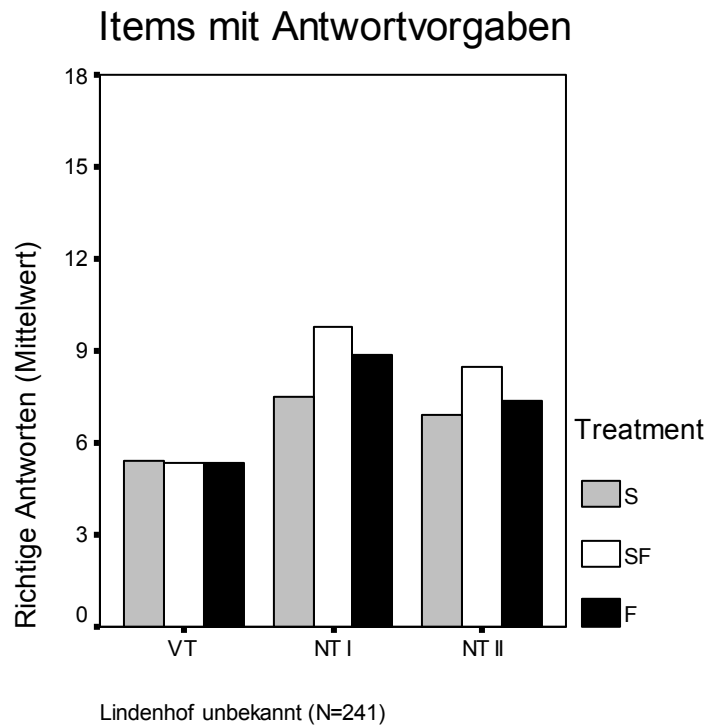


Abbildung 49: Vergleich zwischen den Treatmentgruppen bei den Lindenhof-Nicht-Kennern bei *KogMit18*. Erläuterungen im Text.

Auf der Ebene der einzelnen Treatmentgruppen lassen sich für F- und SF-Schüler Erwerb, Persistenz und Verlust von Wissen nachweisen, für S-Schüler dagegen nur Erwerb und Persistenz von Wissen, nicht aber Vergessen (F-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$; $N = 79$; SF-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$; $N = 81$; S-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(VT < NTI) < 0,001$, $p(VT < NTII) < 0,001$, $p(NTI - NTII) = 0,067$; $N = 81$; vgl. Abbildung 49). Damit hebt sich Treatment S von den beiden übrigen ab.

Die Itemkonstruktion hat Einfluss auf die Ergebnisse der Lindenhof-Nicht-Kenner bei *KogMit18* (vgl. Tabelle 15). SF-Probanden profitieren am meisten von der Itemkonstruktion.

Tabelle 15: Treatmentdifferenzen der Lindenhof-Nicht-Kenner jeweils bezogen auf alle 18 Items bzw. die jeweils sechs F-, SF- bzw. S-Items. Im Vortest gibt es keine statistisch bedeutsamen Differenzen zwischen den Treatmentgruppen. Für die beiden Nachtests sind alle zu untersuchenden Treatmentpaare in der Spaltenbeschriftung aufgeführt. Im Falle statistisch abgesicherter Vorteile einer Treatmentgruppe steht das entsprechende Kürzel im Tabellenfeld. Beispiel SF-Items: NT I / Paarung F-SF → SF, d.h. die SF-Schüler erreichen signifikant höhere Ergebnisse als die F-Schüler. Die entsprechenden Graphiken und konkreten Werte finden sich im Anhang 31.

	VT	NT I			NT II		
		F-SF	F-S	S-SF	F-SF	F-S	S-SF
alle 18 Items	-	SF	F	SF	SF	-	SF
F-Items	-	F	F	SF	-	-	-
SF-Items	-	SF	-	SF	SF	-	SF
S-Items	-	-	-	-	-	-	-

Zusammenfassung zur kognitiven Aussageebene der Lindenhof-Nicht-Kenner bei Items mit Antwortvorgabe (*KogMit18*):

- Die Schüler der Teilgruppe der Lindenhof-Nicht-Kenner lernen, behalten und vergessen.
- Im Gegensatz zu SF- und F-Schülern vergessen S-Schüler nicht.
- SF-Schüler haben den höchsten Lernzuwachs.
- Die Itemkonstruktion hat Einfluss auf die Ergebnisse.

1.2.1.1.2. Ergebnisse der Lindenhof-Kenner für Items mit Antwortvorgabe (*KogMit18*)

Bei der Vorstellung der Resultate der Schüler, die den Lindenhof vor dem Museumsbesuch schon kennen, soll wie folgt verfahren werden. Vorweg werden die Angaben der Schüler vorgestellt, woher ihnen der Lindenhof bekannt ist, dann folgen die Ergebnisse zum Lernerfolg analog zur Teilgruppe der Lindenhof-Nicht-Kenner.

Im Vortest werden die Schüler befragt, ob sie schon einmal am Lindenhof in Bayreuth gewesen sind. Für 121 der Probanden trifft dies zu. In einer Folgefrage wird deren Aktivität am Lindenhof ermittelt, um die Ursachen für unterschiedliche Lernbedingungen dieser Teilgruppe besser beurteilen zu können. Die Schülerangaben werden wie folgt zusammengefasst (vgl. Tabelle 16).

Tabelle 16: Aktivitäten der Lindenhof-Kenner am Umweltschutz-Informationszentrum Lindenhof vor dem Museumsbesuch. Dargestellt ist jeweils die Anzahl der Schüler. N=121.

„Kinderkurs“ (incl. Kindergeburtstag)	Ausstellung (incl. Museum)	Biologischer Kurs	Mehrfachbesuch	Sonstiges	Keine Angabe
32	23	19	7	7	33

Ob und in welcher Intensität Schüler das Naturkundemuseum kennen, lassen diese Daten nur in begrenztem Maße Rückschlüsse zu. Darum erscheint es wichtig, sich zu vergegenwärtigen, dass - wegen der wenig zentralen Lage des Lindenhofs als ehemaliger Aussiedler-Bauernhof - Kinder, die den Lindenhof besuchen, i. d. R. eigens zum Zwecke des Besuches dorthin gebracht werden und ausreichend Zeit haben. Die offenen räumlichen Verhältnisse und die meist freie organisatorische Gestaltung von Aktivitäten am Lindenhof erlauben es den Schülern, einen Großteil der Räumlichkeiten – so auch das Naturkundemuseum - kennen zu lernen. Daneben dient das Museum nicht selten als Kulisse für Kindergeburtstagsfeiern. Daher erscheint es wahrscheinlich, dass Kinder, die schon am Lindenhof waren, dort einige Zeit verbracht haben und das Naturkundemuseum kennen, auch wenn ein Museumsbesuch nicht die Hauptintention ihres Kommens, vielleicht auch nicht die Hauptaktivität am Lindenhof war. Daher ist es erwartbar, dass diese Teilgruppe im Vortest höheres Vorwissen gegenüber der Gesamtgruppe aller Probanden hat.

Einseitige Beeinflussung von treatmentbedingten Ergebnissen aufgrund unterschiedlicher Vorerfahrungen der Treatmentgruppen am Lindenhof können ausgeschlossen werden. Dafür sprechen die Resultate des exakten Fisher-Tests: In 58,9% der Fälle ergibt sich der Chi-Quadrat-Wert rein zufällig (vgl. Anhang 32).

Für die Lindenhof-Kenner stellen sich ebenfalls, sowohl für die Gruppe aller Lindenhof-Kenner, als auch für die einzelnen Treatmentgruppen die Fragen nach Erwerb, Persistenz und Verlust von Wissen (vgl. Tabelle 17). Dabei ist ihr höheres Ausgangsniveau zu beachten. Ihre größere Nähe zum maximalen Ergebnis im Wissenstest lässt weniger Spielraum nach oben zu und könnte sie dadurch schlechter abschneiden lassen als die Lindenhof-Nicht-Kenner.

Diese Vermutung kann anhand der Befunde nicht bestätigt werden (vgl. Tabelle 17). Die Resultate der Lindenhof-Kenner zeugen von Erwerb und Persistenz des Wissens und weisen im Gegensatz zu den Lindenhof-Nicht-Kennern nicht auf Verlust von erworbenen Inhalten nach 40 Tagen hin (Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(\text{VT-NTI}) < 0,001$, $p(\text{VT-NTII}) < 0,001$, $p(\text{NTI-NTII}) = 0,055$; vgl. Anhang 33). Der Unterschied zwischen Nachtest I und II ist nach den in dieser Arbeit verwendeten Konventionen (Signifikanz ab $p < 0,05$) nicht signifikant, ein Vergessen der Lindenhof-Kenner ist somit nicht nachgewiesen, ein β -Fehler jedoch nicht auszuschließen (vgl. Zöfel 2002, S. 65 ff.). Wissenserwerb, Persistenz und kein Vergessen gilt nicht nur für die Gruppe aller Lindenhof-Kenner, sondern auch für die einzelnen Treatmentgruppen der S-, SF- und F-Schüler der Lindenhof-Kenner. Das fehlende Vergessen aller drei Treatmentgruppen spricht für die Richtigkeit dieser Annahme (S-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(\text{VT-NTI}) < 0,001$, $p(\text{VT-NTII}) = 0,002$, $p(\text{NTI-NTII}) = 0,183$; SF-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(\text{VT-NTI}) < 0,001$, $p(\text{VT-NTII}) < 0,001$, $p(\text{NTI-NTII}) = 0,342$; F-Schüler: Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(\text{VT-NTI}) < 0,001$, $p(\text{VT-NTII}) < 0,001$, $p(\text{NTI-NTII}) = 0,329$; vgl. Anhang 34).

Tabelle 17: Übersicht zum Lernerfolg der Lindenhof-Kenner bei Items mit Antwortvorgabe. N=121.

Items mit Antwortvorgabe	
	Lindenhof-Kenner, N=121
Erwerb (VT<NT I)	+
Persistenz (VT<NT II)	+
Verlust (NT I>NTII)	-
	S-Schüler, N=40
Erwerb (VT<NT I)	+
Persistenz (VT<NT II)	+
Verlust (NT I>NTII)	-
	SF-Schüler, N=38
Erwerb (VT<NT I)	+
Persistenz (VT<NT II)	+
Verlust (NT I>NTII)	-
	F-Schüler, N=43
Erwerb (VT<NT I)	+
Persistenz (VT<NT II)	+
Verlust (NT I>NTII)	-

Diese qualitativen Befunde falsifizieren jedoch nicht die Vermutung, die Lindenhof-Kenner nehmen quantitativ weniger Wissen auf und behalten weniger als die Lindenhof-Nicht-Kenner. Diese quantitative Ebene lässt sich durch den Vergleich des Wissenszuwachses testen.

Die Lindenhof-Kenner schneiden bei beiden Werten gleichgut oder besser ab (Tabelle 18). Die Differenzen sind nicht statistisch bedeutsam (Mann-Whitney-U-Test; beide $p(\text{NT I-VT})=0,972$, $p(\text{NT II-VT})=0,102$). Das höhere Vorwissenniveau beeinflusst Aufnahme und Behalten von neuem Wissen nicht negativ. Damit ist die oben formulierte Vermutung zu verwerfen.

Zusammenfassung zur kognitiven Aussageebene der Lindenhof-Kenner bei Items mit Antwortvorgabe (*KogMit18*) (vgl. Gegenüberstellung im Anhang 36):

- Die Bekanntheit des Lindenhofs spielt offensichtlich bei Items mit Antwortvorgabe eine entscheidende Rolle: In wesentlichen Punkten unterscheiden sich die Ergebnisse der Teilgruppen der Lindenhof-Kenner und der Lindenhof-Nicht-Kenner.
- Die Schüler der Teilgruppe der Lindenhof-Kenner lernen und behalten. Im Unterschied zu den Lindenhof-Nicht-Kennern vergessen sie nicht. Das gilt auch für die Treatmentgruppen der S-, SF- und F-Kenner.
- Zwischen den Treatmentgruppen gibt es als weitere Diskrepanz zu den Lindenhof-Nicht-Kennern keine signifikant unterschiedlichen Ergebnisse.
- Die Itemkonstruktion hat - wie auch bei den Lindenhof-Nicht-Kennern - Einfluss auf die Ergebnisse.

1.2.1.2. Items ohne Antwortvorgabe (*KogO*)

Bei den Items mit Antwortvorgaben zeigten sich die Lindenhof-Kenner zu allen Messzeitpunkten überlegen. Dies lässt sich bei den Items ohne Antwortvorgaben nicht bestätigen (vgl. Abbildung 50). Nur im Vortest sind statistische Differenzen nachzuweisen, in den Nachtests I und II verringert sich der Abstand zwischen den Lindenhof-Kennern und den Lindenhof-Nicht-Kennern so weit, dass keine Signifikanzen mehr zu belegen sind (Mann-Whitney-U-Test, Vortest: $p=0,006$, Nachtest I: $p=0,230$, Nachtest II: $p=0,625$).

Die Bekanntheit des Lindenhofs beeinflusst die Testergebnisse, wenn auch in geringerem Maße als bei den Items mit Antwortvorgabe. Lindenhof-Nicht-Kenner und Lindenhof-Kenner werden weiterhin getrennt behandelt.

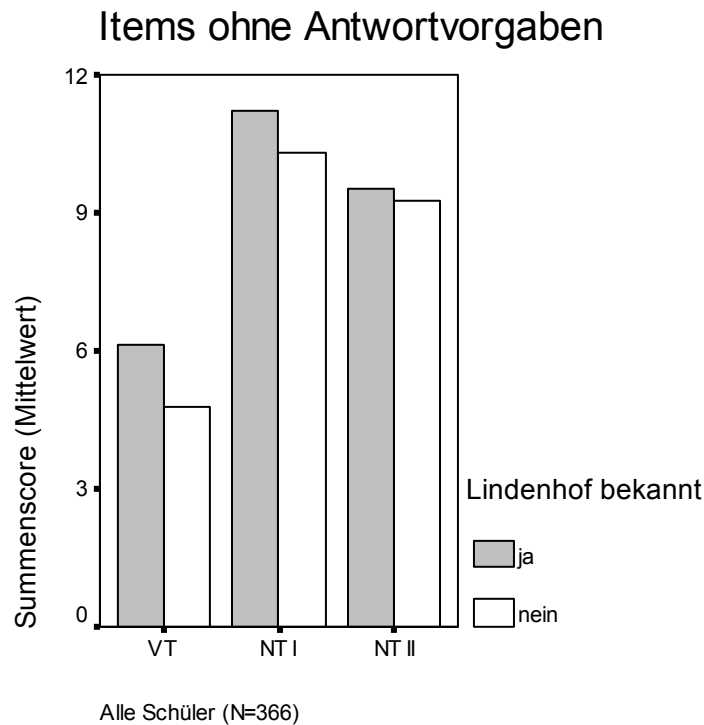


Abbildung 50: Vergleich des Lernerfolgs der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner bei *KogO*. Erläuterungen im Text.

1.2.1.2.1. Ergebnisse der Lindenhof-Nicht-Kenner für Items ohne Antwortvorgabe (*KogO*)

Das Gesamtlernergebnis der Lindenhof-Nicht-Kenner (N=241) weicht nicht von dem der Gesamtgruppe ab: Es gibt Erwerb, Persistenz und Vergessen von Wissen (Friedmann-Test; $p=0,001$; Wilcoxon-Test: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} > NT \text{ II}) = 0,001$; vgl. Anhang 37).

Statistische Unterschiede in den Treatmentgruppen zeigen sich nur im Nachtest I (Kruskal-Wallis-Test; $p=0,037$; vgl. Anhang 38). Wie schon in der Gesamtgruppe bei *KogO* schneiden ebenfalls bei Lindenhof-Nicht-Kennern die S-Schüler besser ab als die SF- und die F-Schüler (Mann-Whitney-U-Test; $p(S > SF) = 0,044$, $p(S > F) = 0,018$; vgl. Anhang 38). Zwischen den Schülern des Treatments SF und F gibt es keine signifikanten Unterschiede (Mann-Whitney-U-Test; $p(SF/F) = 0,665$; vgl. Anhang 38).

Die Analyse der Einzeltreatments ergibt Folgendes: In allen Treatmentgruppen unterscheiden sich die Messwerte (Friedmann-Test; in allen Gruppen $p=0,001$). Während sich bei S- und SF-Schülern Erwerb, Persistenz und Vergessen von Wissen zeigen lassen, kann man bei F-Schülern Erwerb und Persistenz, nicht aber Vergessen finden (Wilcoxon-Test ; S-Schüler: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} > NT \text{ II}) = 0,010$. SF-Schüler: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} > NT \text{ II}) = 0,005$. F-Schüler: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} - NT \text{ II}) = 0,526$; vgl. Anhang 38).

1.2.1.2.2. Ergebnisse der Lindenhof-Kenner für Items ohne Antwortvorgabe (*KogO*)

Die Differenzen zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern scheinen bei den Items ohne Antwortvorgabe aufgehoben zu sein.

Es gibt für Lindenhof-Kenner ebenfalls Erwerb, Persistenz und Vergessen von Wissen (Friedmann-Test; $p=0,001$; Wilcoxon-Test: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} > NT \text{ II}) < 0,001$; vgl. Anhang 39).

Dieser Befund unterscheidet sich von den Ergebnissen der Items mit Antwortvorgaben.

Differenzen zwischen den Treatmentgruppen weist man nur im Nachtest I nach (Kruskal-Wallis-Test; $p=0,013$; vgl. Anhang 40). Die S-Schüler sind wieder besser als die SF- und die F-Schüler (Mann-Whitney-U-Test; $p(S > SF) = 0,007$, $p(S > F) = 0,016$; vgl. Anhang 40). Schüler des Treatments SF und F unterscheiden sich nicht (Mann-Whitney-U-Test; $p(SF/F) = 0,754$; vgl. Anhang 40).

Alle Treatmentgruppen einzeln untersucht, erbringen Unterschiede zwischen den Messzeitpunkten (Friedmann-Test; in allen Gruppen $p < 0,001$). Nur bei S-Schülern ist von Erwerb, Persistenz und Vergessen von Wissen zu sprechen (Wilcoxon-Test; alle zu vergleichenden Paare $p < 0,001$; vgl. Anhang 40), SF- und F-Schüler zeichnen sich durch Erwerb, Persistenz und nicht nachzuweisenden Verlust des Wissens aus (Wilcoxon-Test; SF-Schüler: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} - NT \text{ II}) = 0,191$; F-Schüler: $p(VT < NT \text{ I}) < 0,001$, $p(VT < NT \text{ II}) < 0,001$, $p(NT \text{ I} - NT \text{ II}) = 0,088$; vgl. Anhang 40).

Zusammenfassung zur kognitiven Aussageebene der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner bei Items ohne Antwortvorgabe (*KogO*) (vgl. Gegenüberstellung im Anhang 41):

- Die Ergebnisse der beiden Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner unterscheiden sich bei Items ohne Antwortvorgabe wenig.
- Bei beiden Teilgruppen gibt es Erwerb, Persistenz und Vergessen von Wissen.
- Die S-Schüler haben zwar im Nachtest I bei Lindenhof-Nicht-Kennern und bei Lindenhof-Kennern bessere Ergebnisse, können diesen Vorteil jedoch nicht konservieren. Im Nachtest II fehlen Unterschiede zwischen den Treatmentgruppen.
- Stattdessen zeigen F- (Lindenhof-Nicht-Kenner) bzw. F- und SF-Schüler (Lindenhof-Kenner) nach 40 Tagen keinen nachzuweisenden Wissensverlust.

1.2.1.3. Gegenüberstellung der Hauptergebnisse der Items mit Antwortvorgabe (*KogMit18*) und der Items ohne Antwortvorgabe (*KogO*) differenziert nach der Bekanntheit des Lindenhofs

Für Items mit und ohne Antwortvorgabe gilt für die Gesamtgruppe (N=366) Erwerb, Persistenz und Vergessen von Wissen. Unterschiede in der kognitiven Leistung finden sich zwischen den Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner bei Items mit Antwortvorgabe zu allen Messzeitpunkten: im Vortest, Nachtest I und II; bei Items ohne Antwortvorgabe dagegen nur im Vortest, im Nachtest I und II nicht. Eine Differenzierung in Lindenhof-Nicht-Kenner und Lindenhof-Kenner wurde bei beiden Messverfahren durchgeführt (vgl. Tabelle 20).

Die differenzierte Darstellung der Gegenüberstellung erweist sich als umfangreich: Zwei Ebenen sind zu berücksichtigen, nämlich die Ebene der Messinstrumente (Items mit und ohne Antwortvorgabe, bezeichnet als 1. Ebene) und die Ebene der Teilgruppen (Lindenhof-Nicht-Kenner und Lindenhof-Kenner, bezeichnet als 2. Ebene). Bei dieser Gegenüberstellung interessiert v. a. Gleichheit oder Ungleichheit auf der Ebene der Messinstrumente, demzufolge auf der 1. Ebene. Gleichheit auf 1. Ebene wird erreicht durch gleichsinnige Gleichheit auf 2. Ebene. Beispielsweise kann ein Merkmal bei Items mit Antwortvorgabe / Lindenhof-Nicht-Kenner (+) und bei Items mit Antwortvorgabe / Lindenhof-Kenner (+) zutreffen; gleichzeitig trifft es bei Items ohne Antwortvorgabe / Lindenhof-Nicht-Kenner (+) und bei Items ohne Antwortvorgabe / Lindenhof-Kenner (+) zu; kurz: +/+//+/. Ferner wird Gleichheit auf 1. Ebene erreicht durch gleichsinnige Ungleichheit auf 2. Ebene, z.B. +/-//+/. Ungleichheit auf 1. Ebene wird erreicht durch ungleichsinnige Ungleichheit auf zweiter Ebene, z.B. +/-//+/, durch Gleichheit und Ungleichheit auf 2. Ebene, z.B. +/+//+/-, oder durch ungleichsinnige Gleichheit auf 2. Ebene, z.B. +/-//+/-.

Der kognitive Lernerfolg von Items mit und ohne Antwortvorgabe unabhängig von der Bekanntheit des Lindenhofs besteht in Erwerb und Persistenz von Wissen. Bei Vergessen divergieren die Ergebnisse (Ungleichheit bei Items mit Antwortvorgabe und Gleichheit bei Items ohne Antwortvorgabe auf 2. Ebene → Ungleichheit auf 1. Ebene; +/-//+/-; vgl. Tabelle). Nur bei Items mit Antwortvorgabe/Lindenhof-Kenner wird kein Vergessen belegt.

Teilgruppenanalysen auf der Ebene der Treatments unterstützen dieses Bild: Alle Gruppen zeigen Erwerb und Persistenz von Wissen. Nur bei Vergessen liegen die Unterschiede: S-Schüler vergessen nicht bei Items mit Antwortvorgabe/Lindenhof-Nicht-Kenner und bei Items mit Antwortvorgabe/Lindenhof-Kenner, bei Items ohne Antwortvorgabe vergessen sie in beiden Teilgruppen (ungleichsinnige Gleichheit auf 2. Ebene → Ungleichheit auf 1. Ebene; -/-//+/-; vgl. Tabelle). Sind SF-Schüler Lindenhof-Nicht-Kenner, so zeigen sie Vergessen bei Items mit und ohne Antwortvorgabe, vergessen dagegen nicht als Lindenhof-Kenner bei Items mit und ohne Antwortvorgabe (gleichsinnige Ungleichheit auf 2. Ebene → Gleichheit auf 1. Ebene; +/-//+/-; vgl. Tabelle). F-Schüler vergessen nur als Lindenhof-Nicht-Kenner/Items mit Antwortvorgabe. In allen anderen Fällen vergessen sie nicht (Ungleichheit und Gleichheit auf 2. Ebene → Ungleichheit auf 1. Ebene; +/-//+/-; vgl. Tabelle 20).

Die größten Unterschiede zwischen Items mit und ohne Antwortvorgabe zeigen sich bei der Untersuchung der Treatmentunterschiede. Bei Items mit Antwortvorgabe finden sich bei Lindenhof-Nicht-Kennern in Nachtest I und II Treatmentunterschiede, bei Lindenhof-Kennern sind keine nachzuweisen. Bei Items ohne Antwortvorgabe dagegen finden sich nur im Nachtest I Treatmentunterschiede, diese dafür bei Lindenhof-Nicht-Kennern und bei Lindenhof-Kennern. Der einzige Vergleich von Treatmentunterschieden zwischen Items mit Antwortvorgabe und Items ohne Antwortvorgabe ist im Nachtest I bei Lindenhof-Nicht-Kenner möglich. Items mit Antwortvorgabe ergeben SF-Schüler als beste, S-Schüler als schlechteste Lerner, Items ohne Antwortvorgabe dagegen S-Schüler als die besten Lerner, SF- und F-Schüler unterscheiden sich nicht (vgl. Tabelle). Trotz dieser Differenzen in den Resultaten zu den Treatmentunterschieden lassen sich überraschend viele Gemeinsamkeiten zwischen Items mit und ohne Antwortvorgabe ermitteln.

Tabelle 20: Übersicht über die Gegenüberstellung der Items mit und ohne Antwortvorgabe mit bilanzierender Wertung (1. und 2. Ebene = 1E und 2E) . „+“=“trifft zu“, „-“=“trifft nicht zu“, „G“=“gleich bzw. in der Kernaussage kein Unterschied“, „U“=“ungleich bzw. qualitativer Unterschied in der Aussage“, übrige Abkürzungen wie gehabt. Weiter Erläuterungen im Text.

	Items mit Antwortvorgabe			Items ohne Antwortvorgabe			
	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz 2E	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz 2E	Bilanz 1E
Lernerfolg							
Erwerb	+	+	G	+	+	G	G
Persistenz	+	+	G	+	+	G	G
Verlust	+	-	U	+	+	G	U
	S, N=81	S, N=40		S, N=81	S, N=40		
Erwerb	+	+	G	+	+	G	G
Persistenz	+	+	G	+	+	G	G
Verlust	-	-	G	+	+	G	U
	SF, N=81	SF, N=38		SF, N=81	SF, N=38		
Erwerb	+	+	G	+	+	G	G
Persistenz	+	+	G	+	+	G	G
Verlust	+	-	U	+	-	U	G
	F, N=79	F, N=43		F, N=79	F, N=43		
Erwerb	+	+	G	+	+	G	G
Persistenz	+	+	G	+	+	G	G
Verlust	+	-	U	-	-	G	U
	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz 2E	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz 2E	Bilanz 1E
Treatment-unterschiede							
VT	Keine Unterschiede	Keine Unterschiede	G	Keine Unterschiede	Keine Unterschiede	G	G
NT I	SF>F>S	Keine Unterschiede	U	S>SF=F	S>SF=F	G	U
NT II	SF>F=S	Keine Unterschiede	U	Keine Unterschiede	Keine Unterschiede	G	U

Zusammenfassung zur kognitiven Aussageebene der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner bei Items mit (*KogMit18*) und ohne (*KogO*) Antwortvorgabe (vgl. Gegenüberstellung Tabelle 20):

- Während bezüglich der Lernergebnisse der Gesamtgruppe die Resultate der Items mit Antwortvorgabe auf qualitativer Ebene nicht von denen der Items ohne Antwortvorgabe abweichen, fällt die Analyse bei der Unterscheidung von Lindenhof-Nicht-Kennern und Lindenhof-Kennern unterschiedlicher aus: Beide Teilgruppen erwerben und behalten Wissen; bei Vergessen liegen Unterschiede. Das gilt ebenfalls für alle Typen von Treatmentgruppen (S-, SF-, F-Schüler bei Items mit und ohne Antwortvorgabe, als Lindenhof-Nicht-Kenner und Lindenhof-Kenner). Details siehe Tabelle 20.
- Man findet deutliche Treatmentunterschiede: Bei Items mit Antwortvorgabe sind es die S-Schüler, die etwas schlechtere Ergebnisse haben, dafür aber nicht vergessen. Items ohne Antwortvorgabe liefern gegenläufige Befunde: Hier sind es die F-Schüler, deren Resultate ein etwas niedrigeres Niveau haben, die jedoch nicht vergessen. Details siehe Tabelle 20.

1.2.2. Geschlechtsspezifische kognitive Unterschiede

Vielfach finden sich in den Ergebnissen kognitiver Leistungstests Geschlechtsunterschiede (z.B. Füller 1992, Leibold 1997, Bögeholz 1999, Thompson & Mintzes 2002). Inwieweit sich diese unabhängige Variable in der vorliegenden Untersuchung als bedeutsam erweist, soll geklärt werden.

1.2.2.1. Geschlechtsspezifische kognitive Unterschiede von *KogMit18*

Bei Items mit Antwortvorgabe unterscheiden sich die Wissensergebnisse von Jungen und Mädchen der Gesamtgruppe (N=366) im Vortest, sowie in Nachtest I und II (Mann-Whitney-U-Test, Vortest: $p=0,003$, Nachtest I: $p=0,005$, Nachtest II: $p=0,020$, vgl. Abbildung 51). Die Jungen zeigen die besseren Resultate.

Wie Abbildung 52 zeigt, ist es bei der Teilgruppe der Lindenhof-Nicht-Kenner genauso (Mann-Whitney-U-Test, Vortest: $p=0,008$, Nachtest I: $p=0,006$, Nachtest II: $p=0,007$). Die Jungen schneiden jeweils besser ab. Bei den Lindenhof-Kennern gibt es diese Differenz zu keinem Zeitpunkt (Mann-Whitney-U-Test, Vortest: $p=0,259$, Nachtest I: $p=0,352$, Nachtest II: $p=0,878$; vgl. Abbildung 53). Ist der Lindenhof vorher bekannt, sind durch Items mit Antwortvorgabe keine geschlechtsspezifischen Leistungsunterschiede zu messen.

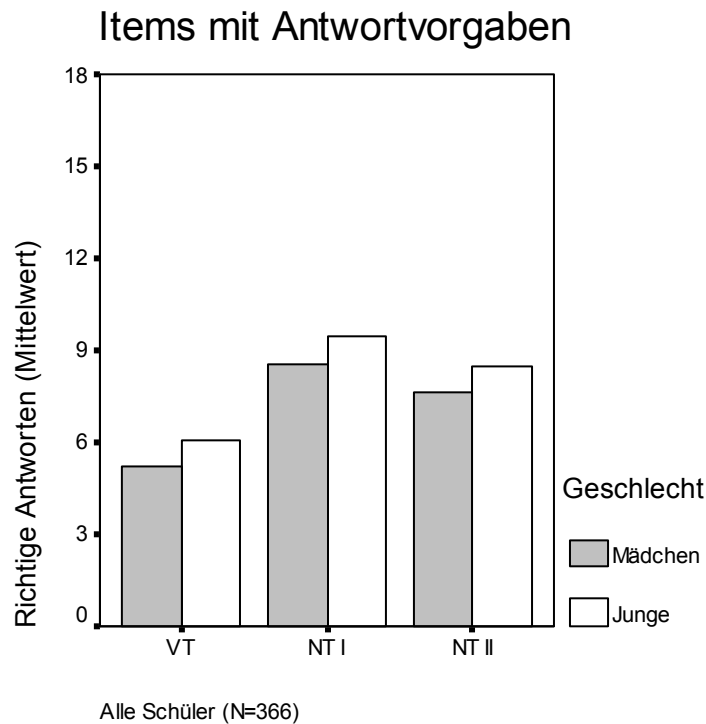


Abbildung 51: Geschlechtsunterschiede im Lernerfolg aller Probanden bei *KogMit18*. Erläuterungen im Text.

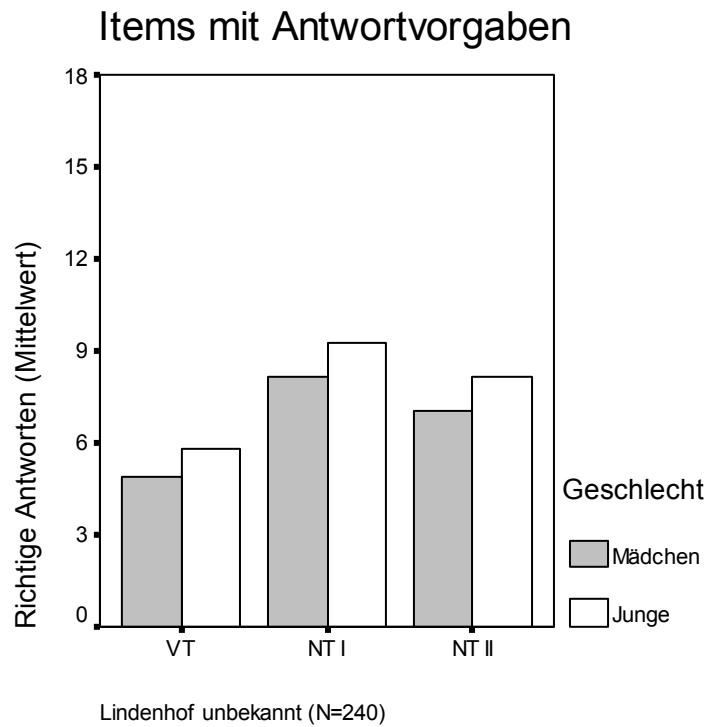


Abbildung 52: Geschlechtsunterschiede im Lernerfolg der Lindenhof-Nicht-Kenner bei *KogMit18*. Erläuterungen im Text.

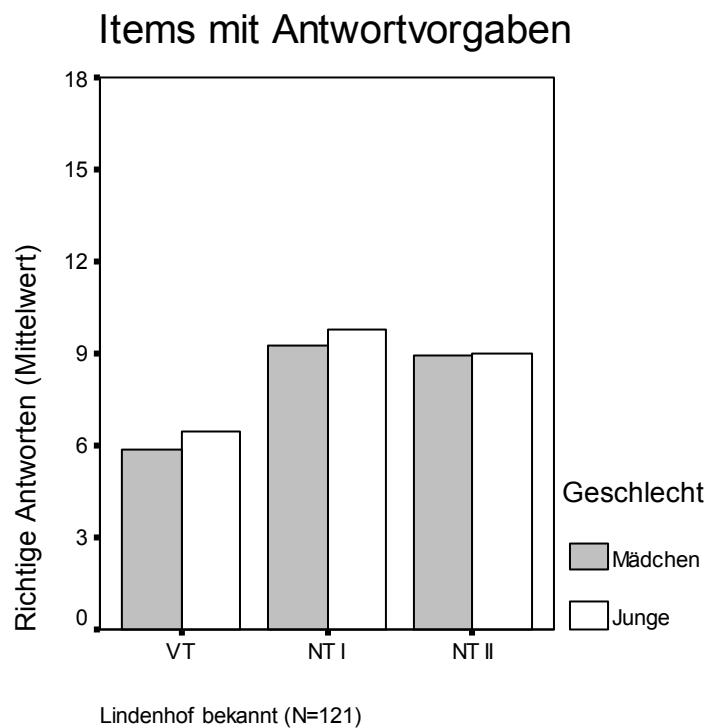


Abbildung 53: Geschlechtsunterschiede im Lernerfolg der Lindenhof-Kenner bei *KogMit18*. Erläuterungen im Text.

Folgende Resultate zu den Treatmentunterschieden sind für eine Gruppierung bezüglich geschlechtsspezifischer Unterschiede bemerkenswert: Bei Jungen und Mädchen lassen sich in Nachtests Treatmentunterschiede nachweisen, niemals zugunsten des selbstbestimmten Treatments, sehr oft zugunsten der Mischform (SF). Das gilt für die Gesamtgruppe und die Gruppe der Lindenhof-Nicht-Kenner. Bemerkenswert in beiden Gruppierungen ist das Fehlen von Treatmentunterschieden bei Mädchen in Nachtest II. Die Lindenhof-Kenner zeigen keine signifikanten Treatmentunterschiede, nicht bei Jungen und nicht bei Mädchen (vgl. Anhang 42).

1.2.2.2. Geschlechtsspezifische kognitive Unterschiede von *KogO*

Items ohne Antwortvorgabe bieten ein ganz anderes Bild. Die Ergebnisse der Gesamtgruppe aller Schüler und der Teilgruppen, Lindenhof-Nicht-Kenner und Lindenhof-Kenner, liefern qualitativ identische Resultate: Zu allen Messzeitpunkten unterscheiden sich Jungen und Mädchen. Immer zeigen Mädchen die besseren Ergebnisse (vgl. Abbildung 54 bis 56). Der Unterschied ist bei den Lindenhof-Kennern sogar etwas größer als bei den Lindenhof-Nicht-Kennern.

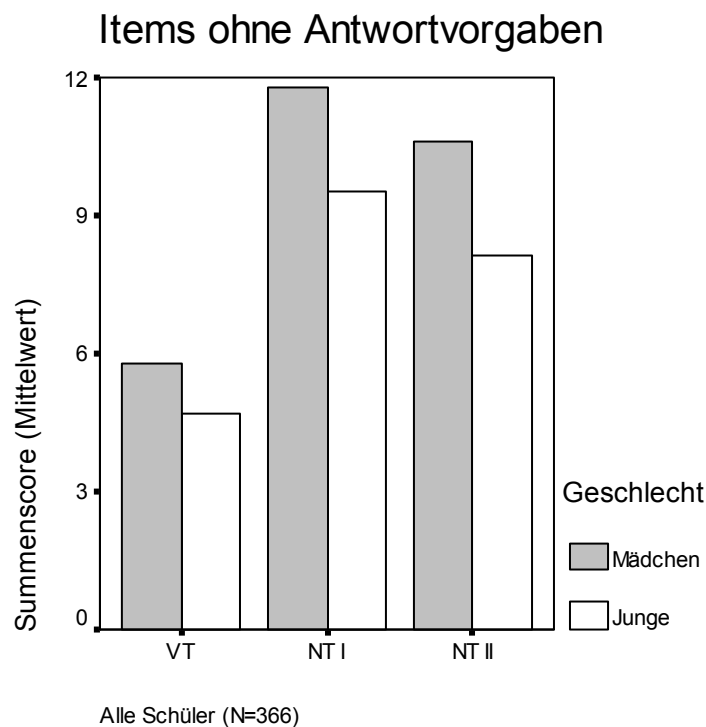


Abbildung 54: Geschlechtsunterschiede im Lernerfolg aller Probanden bei *KogO* (Mann-Whitney-U-Test, Vortest: $p=0,003$, Nachtest I: $p=0,001$, Nachtest II: $p<0,001$; $N=366$).

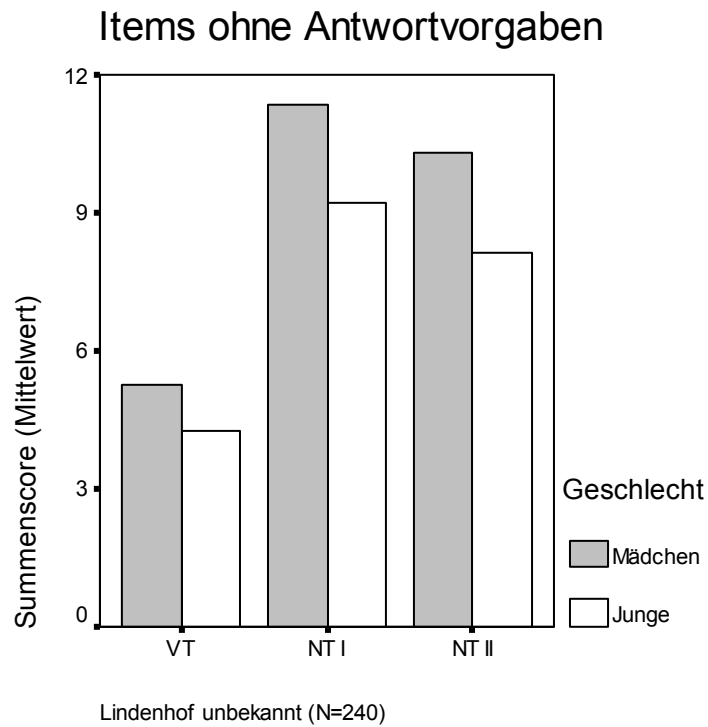


Abbildung 55: Geschlechtsunterschiede im Lernerfolg der Lindenhof-Nicht-Kenner bei *KogO* (Mann-Whitney-U-Test, Vortest: $p=0,032$, Nachtest I: $p=0,008$, Nachtest II: $p=0,006$; $N=240$).

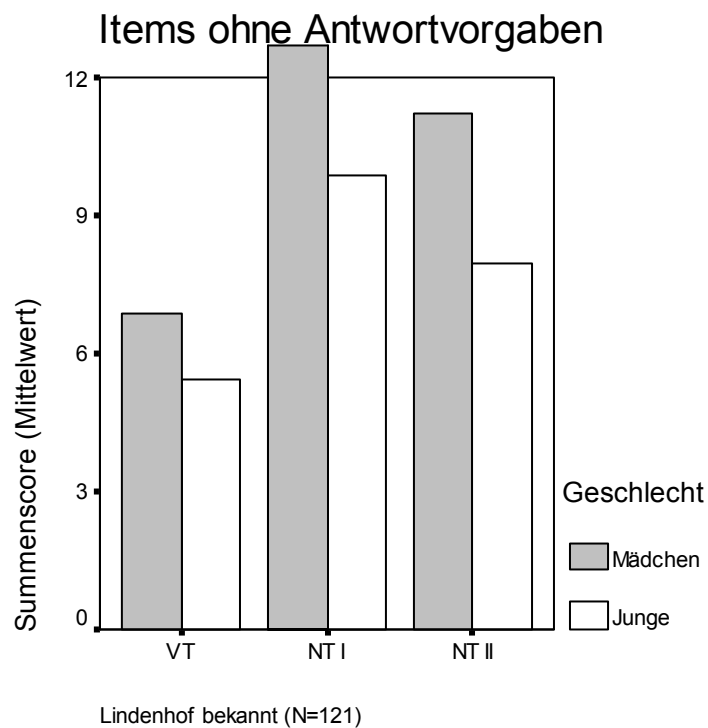


Abbildung 56: Geschlechtsunterschiede im Lernerfolg der Lindenhof-Kenner bei *KogO* (Mann-Whitney-U-Test, Vortest: $p=0,028$, Nachtest I: $p=0,014$, Nachtest II: $p=0,003$; $N=121$).

Treatmentunterschiede lassen sich in diesem Testteil nur in der Gesamtgruppe, nur im Nachtest I und hier nur bei Mädchen nachweisen (Vortest: Kruskal-Wallis-Test; $p=0,286$; Nachtest I: Kruskal-Wallis-Test; $p=0,035$; Nachtest II: Kruskal-Wallis-Test; $p=0,206$; $N=177$). Die Testresultate der S-Mädchen heben sich gegenüber den übrigen positiv ab: Mann-Whitney-U-Test, Nachtest I: $p(S>SF)0,044$, $p(S>F)=0,018$, $p(SF/F)=0,692$.

Zusammenfassung zu geschlechtsspezifischen Unterschieden der Gesamtgruppe und der Teilgruppen, Lindenhof-Nicht-Kenner und Lindenhof-Kenner, bei Items mit und ohne Antwortvorgabe:

- In der Gesamtgruppe unterscheiden sich die Ergebnisse der beiden Messmethoden: Bei Items mit Antwortvorgabe erzielen die Jungen bessere Ergebnisse, bei Items ohne Antwortvorgabe die Mädchen.
- Bei Items mit Antwortvorgabe unterscheiden sich Lindenhof-Nicht-Kenner und Lindenhof-Kenner bezüglich leistungsbezogener Geschlechtsunterschiede: Bei den Nicht-Kennern gibt es Unterschiede zugunsten der Jungen. Bei den Kennern finden sich zwischen den Geschlechtern keine statistischen Unterschiede.
- Bei Items ohne Antwortvorgabe unterscheiden sich, bezogen auf die Tatsache, ob es leistungsbezogene Geschlechtsunterschiede gibt, Lindenhof-Nicht-Kenner und Lindenhof-Kenner **nicht**: In beiden Teilgruppen (Nicht-Kenner und Kenner) sind die Mädchen besser.
- Treatmentunterschiede lassen sich bei Items mit Antwortvorgabe in der Gesamtgruppe und bei Lindenhof-Nicht-Kennern zeigen. Oft zeichnen sich die SF-Schüler durch bessere Ergebnisse aus, niemals die S-Schüler. Bei Mädchen sind im Nachtest II in keiner Gruppierung Treatmentunterschiede nachzuweisen.
- Bei Items ohne Antwortvorgabe gibt es nur in der Gesamtgruppe im Nachtest I der Mädchen Treatmentunterschiede. Das S-Treatment liefert die besten Resultate.

1.3. Die Befunde zur kognitiven Aussageebene im Lichte ihrer Effektstärken

In der Hauptsache interessiert in diesem Abschnitt die Lernleistung der Schüler. Wie Ergebnisse aus den Testteilen *KogMit18* und *KogO* zeigen, verbessern die Schüler ihre Resultate ausgehend von den Vortests zu den Nachtests (vgl. Abbildung 40 und 45). Der Museumsbesuch selbst erklärt mit einer Effektstärke von $\omega^2(KogMit18)=0,57$ (berechnet nach Wolf 2001; S. 99: univariates ω^2 ergibt sich aus dem Prüfwert t des t -Tests; $p<0,001$) bzw. $\omega^2(KogO)=0,53$ (analoge Berechnung; $p<0,001$) mehr als die Hälfte des Lernerfolgs (Interpretation vgl. Häußler et al. 1998, S. 151). Bemerkenswert ist die gute Übereinstimmung der Werte der beiden unterschiedlichen Messmethoden bezüglich der Intervention „Museumsbesuch“. Zu den weiteren Einflussfaktoren zählen Treatment, Geschlecht und Bekanntheit des Lindenhofs. Je nach Messzeitpunkt und Testmethode können sie weitere 10% des Lernerfolgs erklären. Tabelle 21 bestätigt, dass in beiden Testteilen (*KogMit18*, *KogO*) kurzfristig (in Nachtest I) das Treatment der wichtigste dieser drei Faktoren ist. Auffällig ist der relativ große Einfluss des Geschlechts auf die Testleistung bei *KogO*.

Tabelle 21: Aufgeklärte Varianz in Prozent des Einflusses von Treatment, Geschlecht und Bekanntheit des Lindenhofs auf die Lernleistung im Nachtest I. Beispiel: $\omega^2=0,062$ entspricht 6,2% aufgeklärter Varianz; berechnet nach Wolf 2001; S. 99: univariates ω^2 ergibt sich aus dem Prüfwert F der ANOVA.

	Treatment	Geschlecht	Bekanntheit
KogMit18	6,2%	2,3%	1,7%
KogO	4,1%	3,9%	0,6%

Im Nachtest II verschärfen sich die Unterschiede zwischen den Testmethoden. Bei *KogO* ist nur der Einfluss des Geschlechts auf die Testleistung stabil, Treatment und Bekanntheit spielen keine Rolle mehr. Wie Tabelle 22 zeigt, nimmt auch bei *KogMit18* der Einfluss des Treatments ab, bleibt aber bei 3% erklärter Lernleistung noch erwähnenswert. Die Bedeutung des Geschlechts sinkt auf 1,3% ab und ist damit schon im Bereich der Irrelevanz. Interessanterweise ist hier die Bekanntheit mit 3,8% ein relativ bedeutsamer Faktor, während er in *KogO* völlig ohne Einfluss auf die Lernleistung ist.

Tabelle 22: Aufgeklärte Varianz in Prozent des Einflusses von Treatment, Geschlecht und Bekanntheit des Lindenhofs auf die Lernleistung im Nachtest II. Beispiel: $\omega^2=0,030$ entspricht 3,0% aufgeklärter Varianz. Berechnungen analog wie oben.

	Treatment	Geschlecht	Bekanntheit
KogMit18	3,0%	1,3%	3,8%
KogO	0,7%	4,3%	0,0%

Zusammenfassung zu Befunden zur kognitiven Aussageebene im Lichte ihrer Effektstärken:

- Im Nachtest I liefert das Treatment das größte Erklärungspotential. Das gilt für beide Testmethoden.
- Nachtest II dagegen verdeutlicht sehr klar die Unterschiede zwischen *KogMit18* und *KogO*: Bei *KogMit18* sind Treatment und Bekanntheit des Lindenhof von einiger Bedeutung, bei *KogO* nur das Geschlecht.

2. Befunde zur affektiven Aussageebene

Im Folgenden werden die Ergebnisse zur affektiven Ebene berichtet (vgl. Tabelle 23).

Tabelle 23: Übersicht zu den affektiven Lernergebnissen der Gesamtgruppe und der Teilgruppen gebildet nach der Bekanntheit des Lindenhofs und nach dem Geschlecht.

Gesamtgruppe (2.1.)	
Items...	
...mit Antwortvorgabe (<i>AffMu</i> , <i>AffMuUr</i> , <i>Aff-MuPla</i>) (2.1.1.)	...ohne Antwortvorgabe (<i>Freie Äußerung: sprachlich, zeichnerisch; KogO-affektiv</i>) (2.1.2.)
<ul style="list-style-type: none"> • Schülerbewertungen (2.1.1.1.) • Treatmentgruppenvergleich (2.1.1.2.) • Zusammenfassung 	<ul style="list-style-type: none"> • Sprache (<i>Freie Äußerung</i>) (2.1.2.1.1.) • Sprache (<i>KogO</i>) (2.1.2.1.2.) • Zeichnung (2.1.2.2.) • Vergleich zwischen Sprache und Zeichnung (2.1.2.3.) • Zusammenfassung

Teilgruppen (2.2.)	
Teilgruppen nach Bekanntheit des Lindenhofs (2.2.1.)	
Items...	
...mit Antwortvorgabe (<i>AffMu</i>) (2.2.1.1.)	...ohne Antwortvorgabe (<i>Freie Äußerung</i>) (2.2.1.2.)
<ul style="list-style-type: none"> Vergleich der Schülerbewertungen zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern Treatmentgruppenvergleich Zusammenfassung 	<ul style="list-style-type: none"> Vergleich zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern: Sprache, Zeichnung Treatmentgruppenvergleich Zusammenfassung
Teilgruppen nach Geschlecht der Probanden (2.2.2.)	
Items...	
...mit Antwortvorgabe (<i>AffMu</i>) (2.2.2.1.)	...ohne Antwortvorgabe (<i>Freie Äußerung</i>) (2.2.2.2.)
<ul style="list-style-type: none"> Vergleich der Schülerbewertungen zwischen Jungen und Mädchen Treatmentgruppenvergleich 	<ul style="list-style-type: none"> Vergleich zwischen Jungen und Mädchen: Sprache, Zeichnung Treatmentgruppenvergleich
Zusammenfassung	

2.1. Resultate der Gesamtgruppe

Zur affektiven Bewertung des Museumsbesuchs werden unterschiedliche Itemtypen eingesetzt: Analog zum kognitiven Bereich lassen sich Items mit und Items ohne Antwortvorgabe unterscheiden.

2.1.1. Items mit Antwortvorgabe der affektiven Aussageebene

Zu den Items mit Antwortvorgabe zählen hier unterschiedlich differenzierte Schülerbewertungen. Die Probanden sollen durch Likert-Skalen zum Ausdruck bringen, wie ihnen der Museumsbesuch insgesamt, die einzelnen Stationen und Einzelelemente der Stationen gefallen haben (im Fragebogenteil *AffMu*, *AffMuUr* und *AffMuPla* in Nachtest I und II). Diese Items sind nur in Nachtest I und II einsetzbar. Daneben wird ein Vortest-Nachtest-Vergleich auf affektiver Ebene versucht (Ergebnisse des Fragebogens „*Tieraffektivität*“ siehe Anhang 19).

2.1.1.1. Schülereinschätzung des Museumsbesuchs

Die Gesamtbewertung des Museumsbesuchs fällt äußerst positiv aus, die Einschätzungen für die Stationen ergeben ein kaum schlechteres Bild. Dabei sind die Werte von Nachtest I mit der Ausnahme von „Archaeopteryx“ signifikant höher als die von Nachtest II (Wilcoxon-Test, mindestens $p < 0,05$; vgl. Tabelle 24).

Tabelle 24: Gefallen des Museumsbesuchs, der sechs Stationen und dessen Abnahme von Nachtest I zu Nachtest II. Bewertung durch Likert-Skala von „außerordentlich gut“ (5) bis „gar nicht [gefallen]“ (1). Dargestellt sind Mittelwert und Standardabweichung. Wilcoxon-Test. N=366.

	NT I	NT II	p
Museum gesamt	4,45±0,74	4,14±0,88	,000
Stationen:			
Urwald	3,84±0,85	3,71±0,84	,002
Wacholderheide	3,78±0,92	3,55±1,06	,000
Sandgrube	3,88±0,85	3,66±0,97	,000
Kulturfolger	3,93±0,96	3,76±1,00	,001
Archaeopteryx	3,93±1,06	3,90±1,05	,384
Platzbedarf	3,90±0,97	3,76±0,98	,012

Die Bewertungen der Einzelelemente der Stationen, die bei „Urwald“ und „Platzbedarf“ erhoben werden, liegen etwa im Bereich der jeweiligen Station (vgl. Tabelle 25).

Tabelle 25: Gefallen von Station und Einzelelementen von Urwald (*AffMuUr*) und Platzbedarf (*AffMuPla*). Dargestellt sind Mittelwerte. Einzelheiten vgl. Anhang 43. N=366.

		NT I	NT II
Urwald			
	Station	3,84	3,71
	Einzelelemente	4,39 bis 3,39	4,13 bis 3,11
Platzbedarf			
	Station	3,90	3,76
	Einzelelemente	4,37 bis 3,37	4,04 bis 3,25

Zwischen der Gesamtbewertung und allen Stationen finden sich positive Korrelationen, so wie zwischen Urwald und Platzbedarf und entsprechenden Einzelementen (vgl. Material und Methoden: 4.2.1. Konstruktion des Fragebogens und Anhang 17 und 18).

2.1.1.2. Vergleich zwischen den Treatmentgruppen

Selbstgesteuertes Lernen sollte - besser als Formen instruktionsbestimmten Unterrichts - intrinsische Motivation wecken können (Cooper, 1993, S. 17) und damit eher positive affektive Ebenen ansprechen. Wie Kruskal-Wallis-Tests belegen, gibt es bei einem Item Unterschiede zwischen den Treatments (vgl. Anhang 44). Es handelt sich um das Item zur Station Platzbedarf im Nachtest II ($p=0,007$). Schüler des F-Treatment erinnern diese Station verglichen mit S- und SF-Schülern als angenehmer (Mann-Whitney-U-Test; $p(S<F)=0,033$, $p(SF<F)=0,002$). Bei zwölf Tests (sechs Items im Fragebogen, zwei Messzeitpunkte) ist dies kein großer Effekt¹².

Treatmentunterschiede bezüglich des Gefallens spielen keine beherrschende Rolle. Zu keinem Zeitpunkt, nicht bezüglich der Gesamtbewertung und nicht bei den Einschätzungen zu den Stationen, gibt es signifikante Unterschiede zugunsten des selbstgesteuerten Treatments.

Zusammenfassung zu Items mit Antwortvorgaben der affektiven Aussageebene der Gesamtgruppe der Probanden:

- Die affektive Gesamtbewertung der Schüler, wie auch ihre Bewertung einzelner Stationen und von Teilen aus Stationen fällt positiv aus.
- Die drei Treatmentgruppen bewerten den Museumsbesuch und die Stationen in etwa gleichgut.

¹² Einer Bonferroni-Korrektur würde er nicht standhalten (vgl. Zöfel 2002, S. 69): Die Signifikanzgrenze würde auf $0,05/n$, hier $0,05/12=0,004$, herabgesetzt und somit deutlich verschärft. Ob eine solche Korrektur hier nötig ist, wäre zu diskutieren. Bei hypothesengeleitetem Vorgehen wie hier nimmt man i.d.R. davon Abstand.

2.1.2. Items ohne Antwortvorgabe der affektiven Aussageebene

Diese Items nehmen im Fragebogen nur geringen Raum ein, vervollständigen jedoch das Gesamtbild. Die Daten werden v. a. mit Hilfe deskriptiver statistischer Methoden aufbereitet. Es gibt im Test Items für sprachliche und zeichnerische Antwortmöglichkeiten.

2.1.2.1. Sprachliche Äußerungen

Zwei Stellen im Fragebogen erlauben entsprechende freie Äußerungen, die auf affektiver Ebene interpretierbar sind. Zum einen das vorletzte Item des Fragebogens (Fragebogenteil *Freie Äußerung*), zum anderen die offenen Items in Fragebogenteil *KogO*.

2.1.2.1.1. Item zur freien Bemerkung (Fragebogenteil *Freie Äußerung*)

Folgende Resultate erbringt das Item „Willst Du zu Deinem Museumsbesuch noch etwas sagen?“. Im Nachtest I geben 30,1% der Schüler zu dieser Frage keinen Kommentar ab, 7,1% äußern sich u. a. kritisch (Schülerbeispiel: „Es war sehr schön, die Aufgaben aber [waren] etwas schwer, aber sonst [gefielen sie mir] gut.“), 62,8 % dagegen ausschließlich positiv (Schülerbeispiel: „Es hat mir sehr gut gefallen. Macht weiter so!“). Im Nachtest II sinkt der Wert positiver Äußerungen zugunsten der übrigen Kategorien um etwa 10%).

Zwischen der freien Bemerkung und den Treatmentgruppen besteht kein Zusammenhang (Exakter Fisher-Test; $p=0,826$), d.h. es gibt keine Treatmentunterschiede.

2.1.2.1.2. Affektive Ebene der Schülerantworten zu Fragebogenteil *KogO*

Im Fragebogenteil *KogO*, in der Hauptfunktion zur Bestimmung kognitiven Lernens, werden von einigen Schülern affektive Ebenen berührt, nämlich bei den Biotopen (Urwald, Sandgrube, Wacholderheide) und bei Kulturfolger. Bei *Archaeopteryx* und

Platzbedarf zusammen gibt es insgesamt nur vier affektive Äußerungen zu allen drei Messzeitpunkten, so dass diese Stationen vernachlässigt werden können. Bei den übrigen gibt es im Mittel weniger als sieben Prozent entsprechender Äußerungen. Eine bilanzierende Gesamtauswertung soll genügen (vgl. Tabelle 26).

Tabelle 26: Darstellung rein affektiver oder affektiv gefärbter Schüleräußerungen im Fragebogenteil C. Die Kategorien „positiv“, „negativ“ und „neutral“ beziehen sich jeweils auf die Färbung der Schüleräußerung, unabhängig von Richtigkeit oder Zielrichtung der Aussage. Beispiel: positiv: „schöner Teich in der Sandgrube“; negativ: „Hasen wurden regelrecht zubetoniert“; neutral: „außergewöhnliche Tiere“.

	VT	NT I	NT II
Σ positiv	62	95	65
Σ negativ	8	18	13
Σ neutral	9	21	15
Σ gesamt	79	134	93

Unabhängig von der Färbung der Aussage steigt die Anzahl affektiver Schüleräußerungen nach dem Museumsbesuch an. Dabei überwiegen deutlich positive Einschätzungen. Diese Ergebnisse könnte man zu Vortest-Nachtest-Vergleichen heranziehen, die - insbesondere da die Ermittlung von anderen Vortestwerten im affektiven Bereich problematisch ist – wertvoll wären. Jedoch der geringe Anteil der Probanden, die sich affektiv interpretierbar äußern, machen detailliertere Analysen überflüssig, wenngleich die Daten in großer Differenziertheit¹³ erfasst wurden. Aus Gründen der Darstellbarkeit werden sie hier zu drei Kategorien reduziert. Fragebogenteil *KogO* ist in erster Linie zur Ermittlung des kognitiven Lernerfolgs zu verwenden.

2.1.2.2. Zeichnerische Äußerungen

Etwa 65% der Schüler liefern im Nachtest I auswertbare Zeichnungen ihres Lieblingsinhalts. Nur etwa 20% zeichnen gar nicht, 15% Objekte, die offensichtlich nichts mit dem Museumsbesuch zu tun haben oder die zu undeutlich gezeichnet sind.

¹³ Zunächst wurden für jedes Item sechs affektive Kategorien gesucht und anhand von Musteraussagen aufgrund der Einschätzung von vier Experten festgelegt. Die Schüleraussagen wurden von zwei Signifiern entsprechend eingeordnet. Die geringe Quote affektiver Äußerungen machte diese Differenzierung überflüssig.

Hauptinhalt der meisten Zeichnungen sind Elemente des Kindermuseums, obwohl verglichen mit der Dauer der gesamten Exkursion nur ein zeitlich geringer Anteil dort stattfindet (vgl. Anhang 45). Diese Befunde bestätigen Detailanalysen der zeichnerischen Äußerungen: Von den zehn häufigsten gezeichneten Details zählen sechs ausschließlich zum Kindermuseum, nur eines allein zum Naturkundemuseum, zwei sind in beiden Museumsteilen zu sehen. Ein oft gezeichnetes Detail, auf Platz sechs der Häufigkeitsrangliste, wird an keiner Stelle thematisiert, ein Terrarium mit lebenden Eurasischen Zwergmäusen (*Micromys minutus*, Pallas 1778) (vgl. Anhang 46).

Zusammenhänge zwischen dem durchlaufenen Treatment und dem Hauptinhalt der Zeichnung lassen sich nicht feststellen (Exakter Fisher-Test; Nachtest I: p-Wert nicht berechenbar mit N=291 angefertigte Zeichnungen; Nachtest II: $p=0,502$ bei N=176 angefertigten Zeichnungen; vgl. Anhang 47). Interessanterweise ergeben sich im Nachtest II Zusammenhänge zwischen dem Treatment und der Tatsache, ob gezeichnet wird (Exakter Fisher-Test; Nachtest I: $p=0,431$; Nachtest II: $p=0,043$; N=366 bei beiden Testzeitpunkten; vgl. Anhang 48). Die SF-Probanden zeichnen im Nachtest II überzufällig mehr als man laut Statistik erwarten sollte. Die beiden anderen Treatmentgruppen zeichnen etwa in gleichem Maße weniger.

2.1.2.3. Sprachliche und zeichnerische Äußerungen

Die freien Äußerungen der Schüler ergeben im Nachtest I ein positives Gesamtbild. 63% der Probanden äußern sich sprachlich positiv, 65% liefern auswertbare Zeichnungen ihres Lieblingsinhalts. Zwischen auswertbaren Zeichnungen und sich sprachlich positiv äußernden Schülern gibt es keinen Zusammenhang (Exakter Fisher-Test; $p=0,600$). Vermutlich sprechen die unterschiedlichen Erhebungsinstrumente nicht dieselben Probanden an, sondern ergänzen sich.

Zusammenfassung zu Items ohne Antwortvorgaben der affektiven Aussageebene der Gesamtgruppe der Probanden:

- Über die Hälfte aller Schüler äußern sich in einem fakultativ zu beantwortenden und sehr offen gestalteten Item ausschließlich positiv über den Museumsbesuch.
- Der Fragebogenteil *KogO* ist nur für Aussagen auf kognitiver Ebene zu verwenden und nicht für die affektive Ebene.
- Zeichnen erweist sich als eine Ergänzung zur ausschließlich sprachvermittelten Datenerhebung: Die Schüler zeichnen v.a. Elemente des Kindermuseums.
- Es werden wenige Treatmentunterschiede nachgewiesen. Im Nachtest II zeichnen SF-Schüler überzufällig mehr als zu erwarten.

2.2. Resultate von Teilgruppen

2.2.1. Bekanntheit des Lindenhofs

2.2.1.1. Items mit Antwortvorgabe

Betrachtet werden die Gesamtbewertung des Museumsbesuchs und die Bewertung der einzelnen Stationen unter Berücksichtigung der Bekanntheit des Lindenhofs. Auf beiden Ebenen bewerten beide Teilgruppen den Museumsbesuch positiv (vgl. Anhang 49). In der Gesamtbewertung finden sich zwischen den Teilgruppen, Lindenhof-Kennern und Lindenhof-Nicht-Kennern, im Nachtest II Differenzen zugunsten der Lindenhof-Nicht-Kenner (Mann-Whitney-U-Test; Nachtest I: $p=0,127$; Nachtest II: $p=0,017$; vgl. Anhang 49). Die Einzelbetrachtung der Stationen erbringt in Nachtest I und II i.d.R. keine Unterschiede zwischen den Teilgruppen. Ein typisches Beispiel stellt „Platzbedarf“ dar (Mann-Whitney-U-Test; Nachtest I: $p=0,253$; Nachtest II: $p=0,749$; vgl. Tabelle 27; zu beachten sind die Zeilen). Die Ausnahme ist „Wacholderheide“: Es gibt Unterschiede im Nachtest I zugunsten der Lindenhof-Nicht-Kenner (Mann-Whitney-U-Test; Nachtest I: $p=0,027$; Nachtest II: $p=0,221$; vgl. Tabelle 28; zu beachten sind die Zeilen). Zwischen Lindenhof-Kennern und Lindenhof-Nicht-Kennern gibt es im direkten Vergleich der Werte wenige statistische Unterschiede.

Tabelle 27: Gegenüberstellung der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner mit Berücksichtigung der Testzeitpunkte der Station „Platzbedarf“.

	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	p
Nachtest I	3,86	3,97	0,253
Nachtest II	3,76	3,75	0,749
p(NT I- NT II)	0,210	0,008	

Tabelle 28: Gegenüberstellung der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner mit Berücksichtigung der Testzeitpunkte der Station „Wacholderheide“.

	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	p
Nachtest I	3,86	3,62	0,027
Nachtest II	3,49	3,65	0,221
p(NT I- NT II)	>0,001	0,917	

Eine Betrachtung der Veränderung der Werte von Nachtest I zu Nachtest II ergibt für die Gesamtbewertung bei beiden Teilgruppen signifikante Unterschiede zwischen Nachtest I und II. Eine analoge Analyse der einzelnen Stationen legt Unterschiede zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern offen (vgl. Anhang 50): In vier von sechs Fällen findet man bei Lindenhof-Nicht-Kennern Unterschiede zwischen Nachtest I und II, bei Lindenhof-Kennern gibt es dagegen in fünf von sechs Fällen keine statistischen Unterschiede zwischen Nachtest I und II. Ein typisches Beispiel hierfür ist die Station „Wacholderheide“ (Wilcoxon-Test; Lindenhof-Nicht-Kenner $p(\text{NT I} > \text{NT II}) < 0,001$; Lindenhof-Kenner $p(\text{NT I} - \text{NT II}) = 0,917$; vgl. Tabelle 28; zu beachten sind die Spalten). Die positive Schülerbewertung der Stationen nimmt offensichtlich bei Lindenhof-Nicht-Kennern leichter ab als bei Lindenhof-Kennern. Eine Ausnahme stellt die Station Platzbedarf dar: Die Verhältnisse sind umgekehrt (Wilcoxon-Test; Lindenhof-Nicht-Kenner $p(\text{NT I} - \text{NT II}) = 0,210$; Lindenhof-Kenner $p(\text{NT I} > \text{NT II}) = 0,008$; vgl. Tabelle 27; zu beachten sind die Spalten).

In der Gesamtgruppe spielen für die Schülerbewertung des Museums die Treatments keine wichtige Rolle. In der Teilgruppe der Lindenhof-Nicht-Kenner gibt es bei zwei Stationen jeweils im Nachtest II differierende Schüleräußerungen bezüglich der Treatments (Kruskal-Wallis-Test; $p(\text{Urwald}) = 0,024$, $p(\text{Platzbedarf}) = 0,017$): Bei „Platzbedarf“ zugunsten von Treatment F im Vergleich mit SF (Mann-Whitney-U-

Test, $p(F/SF)=0,004$), bei „Urwald“ zugunsten von F und SF jeweils im Vergleich mit Treatment S (Mann-Whitney-U-Test, $p(S/SF)=0,021$, $p(S/F)=0,020$; vgl. Anhang 51). In der Gruppe der Lindenhof-Kenner lassen sich dagegen keine statistischen Unterschiede zwischen den Treatments nachweisen (vgl. Anhang 52). Damit spielen für beide Teilgruppen Treatmentunterschiede zum Gefallen des Museumsbesuchs, keine beherrschende Rolle. Zu keinem Zeitpunkt, nicht bezüglich der Gesamtbewertung und nicht bei den Einschätzungen zu den Stationen, gibt es signifikante Unterschiede zugunsten des selbstgesteuerten Treatments.

Zusammenfassung zu den Items mit Antwortvorgabe der affektiven Aussageebene der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner:

- Auf allen Ebenen wird der Museumsbesuch von beiden Teilgruppen positiv bewertet.
- Lindenhof-Nicht-Kenner bewerten den Besuch in einigen Fällen positiver als Lindenhof-Kenner.
- Die Einschätzung der Lindenhof-Kenner nimmt dagegen bei den meisten Stationen vom Nachtest I zum Nachtest II - im Gegensatz zu den Lindenhof-Nicht-Kennern - nicht signifikant ab und ist damit stabiler.
- Treatmentunterschiede spielen eine geringe Rolle. S-Schüler vergeben nie die signifikant positivsten Bewertungen.

2.2.1.2. Items ohne Antwortvorgabe

Bezüglich des ausschließlich sprachvermittelten Items „Willst Du zu Deinem Museumsbesuch noch etwas sagen?“ äußern sich in der Gesamtgruppe im Nachtest I über 60% der Probanden positiv. Die Teilgruppen liefern folgende Resultate: Die positive Einschätzung von Lindenhof-Nicht-Kennern und Lindenhof-Kennern unterscheidet sich. Der Wert der Lindenhof-Kenner liegt im Nachtest I um etwa 11%, im Nachtest II um etwa 7% niedriger als der der Lindenhof-Nicht-Kenner. Statistisch bedeutsam ist der Zusammenhang zwischen Bekanntheit des Lindenhofs und dem Hauptinhalt dieser freien sprachlichen Äußerung nicht (Exakter Fisher-Test; Nachtest I: $p=0,081$; Nachtest II: $p=0,447$ beide mit $N=362$). Ein Zusammenhang zwischen Treatment und Inhalt lässt sich ebenfalls nicht nachweisen (Exakter Fisher-Test; Nachtest I: $p=0,826$; Nachtest II: $p=0,432$ beide mit $N=366$). Insgesamt äußern sich

in beiden Teilgruppen mehr als die Hälfte aller Schüler zu diesem fakultativ zu kommentierenden Item. Die meisten Schüler äußern sich positiv.

Zwischen dem Hauptinhalt der Zeichnung und der Bekanntheit des Lindenhofs gibt es keinen signifikanten Zusammenhang (Exakter Fisher-Test; Nachtest I: $p=0,681$ bei $N=289$ Zeichnungen mit Angabe der Bekanntheit des Lindenhofs; Nachtest II: $p=0,078$ bei $N=174$ Zeichnungen mit Angabe der Bekanntheit des Lindenhofs).

Unterschiede zeigen sich dagegen in der Frage, ob gezeichnet wird: Nur wenige Schüler verweigern sich im Nachtest I der Zeichnung. Die Nicht-Zeichner sind in etwa gleich auf Lindenhof-Nicht-Kenner (ca. 19%) und Lindenhof-Kenner (über 22%) verteilt. Im Nachtest II steigt der Anteil der Nicht-Zeichner bei Lindenhof-Nicht-Kennern auf fast 43% an, bei Lindenhof-Kennern auf etwa 70%. Hier unterscheiden sich die beiden Teilgruppen offenbar (Exakter Fisher-Test; Nachtest I: $p=0,489$; Nachtest II: $p<0,001$ beide mit $N=362$).

In beiden Teilgruppen (Lindenhof-Nicht-Kenner, Lindenhof-Kenner) lassen sich keine Treatmentunterschiede nachweisen weder in den Hauptinhalten noch in der Frage, ob gezeichnet wird (Exakter-Fisher-Test; Lindenhof-Nicht-Kenner/ Hauptinhalt: Nachtest I: $p=0,165$ mit $N=195$; Nachtest II: $p=0,673$ mit $N=138$; Lindenhof-Kenner/ Hauptinhalt: Nachtest I: $p=0,088$ mit $N=94$; Nachtest II: $p=0,897$ mit $N=36$; Lindenhof-Nicht-Kenner/ Zeichnung Ja/Nein: Nachtest I: $p=0,617$; Nachtest II: $p=0,059$ beide $N=241$; Lindenhof-Kenner/ Zeichnung Ja/Nein: Nachtest I: $p=0,311$; Nachtest II: $p=0,798$ beide $N=121$).

Zusammenfassung zu Items ohne Antwortvorgabe der affektiven Aussageebene der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner:

- In sprachvermittelten und v.a. zeichnerischen Ebenen deuten sich positivere Bewertungen bzw. größeres Engagement durch Lindenhof-Nicht-Kenner an.
- Beziehungen zwischen Treatment und Ergebnissen der freien Äußerungen in Zeichnung oder Sprache finden sich in beiden Teilgruppen nicht.

2.2.2. Geschlechtsspezifische affektive Unterschiede

2.2.2.1. Items mit Antwortvorgabe

Hier unterscheiden sich in Nachtest I und II die Einschätzungen von Jungen und Mädchen bezüglich der Gesamtbewertung des Museumsbesuches. Mädchen schätzen den Besuch signifikant positiver ein als Jungen (Mann-Whitney-U-Test; Nachtest I: $p(\text{Mädchen} > \text{Jungen}) = 0,003$; Nachtest II: $p(\text{Mädchen} > \text{Jungen}) < 0,001$; vgl. Anhang 53). Die Einzelergebnisse der Stationen bestätigen dieses Bild nur z.T.: Im Nachtest I und II bewerten Mädchen bis auf „Archaeopteryx“ alle Stationen besser als Jungen. In Nachtest I ist diese Bewertung nur bei Station „Kulturfolger“ signifikant besser, im Nachtest II bei „Kulturfolger“, „Wacholderheide“ und „Platzbedarf“ (Beispiel „Wacholderheide“: Mann-Whitney-U-Test; Nachtest I: $p(\text{Mädchen} / \text{Jungen}) = 0,179$; Nachtest II: $p(\text{Mädchen} > \text{Jungen}) = 0,004$; vgl. Anhang 53). Die höhere Einschätzung der Jungen bei „Archaeopteryx“ ist im Nachtest II statistisch bedeutsam (Mann-Whitney-U-Test; Nachtest I: $p(\text{Mädchen} / \text{Jungen}) = 0,066$; Nachtest II: $p(\text{Mädchen} < \text{Jungen}) < 0,048$; vgl. Anhang 53).

Interessanterweise unterscheiden sich Lindenhof-Nicht-Kenner und Lindenhof-Kenner durch unterschiedliche Bewertungen der Geschlechter. Bei Lindenhof-Nicht-Kennern liegen die Werte der Mädchen in der Gesamtbewertung in Nachtest I und II positiver als die der Jungen (Mann-Whitney-U-Test; Nachtest I: $p(\text{Mädchen} > \text{Jungen}) = 0,002$; Nachtest II: $p(\text{Mädchen} > \text{Jungen}) < 0,001$; vgl. Anhang 54). Ähnliches gilt für drei von sechs Stationen im Nachtest I und vier von sechs Stationen im Nachtest II. Beispielhaft sei die Station „Wacholderheide“ aufgeführt (Mann-Whitney-U-Test; Nachtest I: $p(\text{Mädchen} > \text{Jungen}) = 0,009$; Nachtest II: $p(\text{Mädchen} > \text{Jungen}) = 0,001$; vgl. Anhang 54). Im Gegensatz dazu finden sich bei Lindenhof-Kennern in der Gesamtbewertung zwischen den Geschlechtern in Nachtest I und II keine Unterschiede. Die Unterschiede in den Stationen gleichen sich aus: Bei „Kulturfolger“ äußern sich in Nachtest I und II Mädchen positiver, bei „Archaeopteryx“ Jungen (Mann-Whitney-U-Test; „Kulturfolger“: Nachtest I: $p(\text{Mädchen} > \text{Jungen}) = 0,023$; Nachtest II: $p(\text{Mädchen} > \text{Jungen}) = 0,013$; „Archaeopteryx“: Nachtest I: $p(\text{Mädchen} < \text{Jungen}) = 0,007$; Nachtest II: $p(\text{Mädchen} < \text{Jungen}) = 0,001$; vgl. Anhang 55).

Trotz weniger Treatmentunterschiede deuten sich Differenzen zwischen Jungen und Mädchen an. Wenn Jungen einer Treatmentgruppe Stationen positiver bewerten als die übrigen, so sind es immer die F-Schüler. In der Gesamtgruppe zeigen sich in drei Stationen signifikante Ergebnisse (Kruskal-Wallis-Tests; Nachtest I: „Wacholderheide“: $p=0,013$, „Archaeopteryx“: $p=0,046$; Nachtest II: „Platzbedarf“: $p=0,012$), von den sechs Paarvergleichen, an denen das F-Treatment beteiligt ist, fallen fünf statistisch bedeutsam zugunsten der F-Jungen aus (Mann-Whitney-U-Test; Nachtest I: „Wacholderheide“: $p(S/SF)=0,273$, $p(S<F)=0,006$, $p(F>SF)=0,037$, „Archaeopteryx“: $p(S/SF)=0,265$, $p(S/F)=0,224$, $p(F>SF)=0,011$; Nachtest II: „Platzbedarf“: $p(S/SF)=0,459$, $p(S<F)=0,031$, $p(F>SF)=0,005$; vgl. Anhang 56). Bei Lindenhof-Nicht-Kennern finden sich noch bei einer Station besonders positive Urteile durch F-Jungen (Lindenhof-Nicht-Kenner: Kruskal-Wallis-Test; Nachtest II: „Platzbedarf“: $p=0,045$; Mann-Whitney-U-Test; Nachtest II: „Platzbedarf“: $p(S/SF)=0,209$, $p(S/F)=0,229$, $p(F>SF)=0,014$; vgl. Anhang 56). Bei Lindenhof-Kennern gibt es keine Treatmentunterschiede (vgl. Anhang 56). Bei Mädchen ist die Bewertung differenzierter. In der Gesamtgruppe lassen sich gar keine Bewertungsunterschiede feststellen. In der Gruppe der Lindenhof-Nicht-Kenner gibt es in zwei Stationen signifikante Unterschiede (Lindenhof-Nicht-Kenner: Kruskal-Wallis-Test; Nachtest II: „Urwald“: $p=0,017$, „Wacholderheide“: $p=0,012$; Mann-Whitney-U-Test; Nachtest II: „Urwald“: $p(S<SF)=0,018$, $p(S<F)=0,036$, $p(F/SF)=0,520$; „Wacholderheide“: $p(S<SF)=0,011$, $p(S/F)=0,756$, $p(F<SF)=0,008$; vgl. Anhang 57). Wie sich zeigt, sind die Urteile der F-Mädchen, öfter die der SF-Mädchen besonders positiv. Nimmt man die Jungen als Vergleichsmaßstab verschiebt sich dieses Bild bei Lindenhof-Kennern noch deutlicher: Bei „Kenner-Mädchen“ heben sich, wenn es Unterschiede gibt, die S-Mädchen durch positive Bewertungen ab (Lindenhof-Kenner: Kruskal-Wallis-Test; „Kulturfolger“: Nachtest I: $p=0,010$, Nachtest II: $p=0,020$; Mann-Whitney-U-Test; „Kulturfolger“: Nachtest I: $p(S>SF)=0,006$, $p(S>F)=0,009$, $p(F/SF)=0,865$; Nachtest II: $p(S>SF)=0,005$, $p(S/F)=0,098$, $p(F/SF)=0,268$; vgl. Anhang 57).

Diese Treatmentunterschiede sind nicht sehr dominant, berücksichtigt man Design und Messinstrument (sechs Items im Fragebogen, zwei Messzeitpunkte). Eine Bonferroni-Korrektur würde die Signifikanzgrenze auf $0,05/n$, hier $0,05/12=0,004$ herabsetzen. Das sich aus den wenigen Unterschieden abzeichnende Bild ist jedoch in sich konsistent und hat Erklärungswert. Folgendes deutet sich an: Ist der Lindenhof

unbekannt, so werden eher kleinschrittig anleitende Aufgaben als positiv angesehen (Jungen / Nicht-Kenner: F; Mädchen / Nicht-Kenner: F oder SF). Ist der Lindenhof dagegen vorher bekannt, so beurteilen die fremdbestimmten Schüler den Museumsbesuch nicht positiver als eher selbstbestimmte Gruppen (Jungen / Kenner: keine Unterschiede; Mädchen / Kenner: S). Anscheinend bevorzugen Mädchen weniger Anleitung als Jungen.

2.2.2.2. Items ohne Antwortvorgabe

Zu analysieren sind Unterschiede in Sprache und Zeichnung. Es gibt in Nachtest I und II Zusammenhänge zwischen der freien Bemerkung und dem Geschlecht der Schüler (Exakter Fisher-Test; Nachtest I und II: $p < 0,001$). Mädchen äußern sich häufiger positiv und seltener negativ als Jungen. Auch zwischen dem Hauptinhalt der Zeichnung und dem Geschlecht sind in Nachtest I und II Unterschiede nachzuweisen (Exakter Fisher-Test; Nachtest I: $p = 0,027$; Nachtest II: $p = 0,006$). Jungen zeichnen z.B. öfter Inhalte, die nichts mit dem Lindenhof zu tun haben. Signifikante Zusammenhänge zwischen Geschlecht und der Tatsache, ob Zeichnungen angefertigt werden, sind nicht zu finden (Exakter Fisher-Test; Nachtest I: $p = 0,120$; Nachtest II: $p = 0,143$).

Bezüglich der Gesamtgruppe ist in Nachtest I und II zu Sprache und Zeichnung nur eine Abweichung von zufälliger Verteilung der Treatmentgruppen zu entdecken. Lediglich im Nachtest II zeichnen SF-Jungen überzufällig häufig (vgl. Anhang 58). Ähnlich selten sind Treatmentunterschiede der Teilgruppen Lindenhof-Nicht-Kenner und Lindenhof-Kenner. Bedeutsame Treatmentunterschiede sind nicht zu erkennen (vgl. Anhang 58).

Zusammenfassung zu Aussagen zur affektiven Aussageebene zu Geschlechtsunterschieden:

- Bei Items mit Antwortvorgabe bewerten Mädchen den Museumsbesuch positiver als Jungen. Genauere Analysen zeigen, dass sich diese Aussage auf Mädchen, die den Lindenhof vor dem Museumsbesuch nicht kennen, beschränkt.
- Die wenigen bei Items mit Antwortvorgabe gefundenen Unterschiede in den Treatmentgruppen deuten folgendes Bild an: F-Jungen geben positivere Urteile über den Museumsbesuch ab als die übrigen Jungen. SF-Mädchen, denen der Lindenhof unbekannt ist, beurteilen den Besuch besonders positiv so wie S-Mädchen, denen der Lindenhof bekannt ist.
- In den Items ohne Antwortvorgabe zu Zeichnung und Sprache unterscheiden sich die Einschätzungen von Jungen und Mädchen ebenfalls. Jungen äußern sich sprachlich und zeichnerisch eher seltener, geben leichter negative Haltungen zu erkennen oder weichen vom Thema des Museumsbesuchs ab.
- Es lassen sich bezüglich der Items ohne Antwortvorgabe kaum Treatmentunterschiede nachweisen.

2.3. Die Befunde zur affektiven Aussageebene im Lichte ihrer Effektstärken (beschränkt auf die Testebene der Items mit Antwortvorgaben *AffMu*)

Der Einfluss des Museumsbesuches selbst ist aufgrund des fehlenden Vortestwertes nicht zu berechnen. Welchen Einfluss Treatment, Geschlecht und Bekanntheit des Lindenhofs auf das Gefallen des Gesamtbesuches und die sechs Stationen haben, sei im Folgenden ausgeführt:

Im Nachtest I ist nur das Geschlecht ein Faktor, der z. T. von einiger Bedeutung ist. Treatment und Bekanntheit des Lindenhofs spielen eine sehr geringe Rolle (vgl. Tabelle 29).

Tabelle 29: Aufgeklärte Varianz der affektiven Einschätzung in Prozent des Einflusses von Treatment, Geschlecht und Bekanntheit des Lindenhofs auf die Lernleistung im Nachtest I. Beispiel: $\omega^2=0,023$ entspricht 2,3% aufgeklärter Varianz. Berechnungen analog zu Ergebniskapitel 1.3.

	Treatment	Geschlecht	Bekanntheit
Museum gesamt	0,1%	2,3%	0,9%
Stationen:			
Urwald	0,7%	0,0%	0,6%
Wacholderheide	0,8%	0,6%	1,5%
Sandgrube	0,3%	0,2%	0,1%
Kulturfolger	1,2%	3,3%	0,0%
Archaeopteryx	0,2%	0,4%	0,4 %
Platzbedarf	0,9%	0,9%	0,3%

Im Nachtest II verschiebt sich das Bild ein wenig. Während in der Gesamtbewertung wiederum nur das Geschlecht entscheidende Bedeutung zur Erklärung der Gesamtvarianz hat, ist auf der Ebene der Treatments zumindest bei „Platzbedarf“ mit 2,9% von einiger Bedeutung für die Erklärung des affektiven Eindruckes der Probanden zu sprechen. Die Bekanntheit des Lindenhofs ist auf beiden Ebenen kaum bedeutsam (vgl. Tabelle 30).

Tabelle 30: Aufgeklärte Varianz der affektiven Einschätzung in Prozent des Einflusses von Treatment, Geschlecht und Bekanntheit des Lindenhofs auf die Lernleistung im Nachtest II. Beispiel: $\omega^2=0,060$ entspricht 6,0% aufgeklärter Varianz. Berechnungen analog wie oben.

	Treatment	Geschlecht	Bekanntheit
Museum gesamt	0,0	6,0	0,6
Stationen:			
Urwald	1,0	1,4	0,2
Wacholderheide	0,3	3,1	0,5
Sandgrube	0,2	0,6	0,0
Kulturfolger	1,0	6,0	0,1
Archaeopteryx	0,7	0,8	0,0
Platzbedarf	2,9	2,2	0,0

Zusammenfassung zu Befunden zur affektiven Aussageebene im Lichte ihrer Effektstärken:

- Treatment und Bekanntheit des Lindenhofs spielen kurz-, wie auch längerfristig kaum eine Rolle.
- Lediglich das Geschlecht ist zur Erklärung v. a. der längerfristigen Einschätzungen ein nicht unwichtiger Punkt.

3. Kognitive und affektive Ergebnisse zum Kindermuseum

Neben den zwei etwa dreißigminütigen Arbeitsphasen im Naturkundemuseum waren die Schüler ebenfalls für dreißig Minuten mit ihrer Betreuerin im Kindermuseum. Alle Treatmentgruppen wurden in diesem Museumsteil identisch behandelt: Ratespiel, Kletterbaum, Kameraspiel (vgl. Material und Methoden 7).

Die Befunde zum kognitiven Lernerfolg beziehen sich ausschließlich auf drei Items, die Teil des erweiterten Fragebogenteils *KogMit25* sind. Bezüglich dieser drei Items kann man für die Gesamtgruppe und die Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner Erwerb, Persistenz und fehlenden Verlust des Wissens feststellen (Gesamtgruppe (N=366): Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(VT < NTI) < 0,001$, $p(VT < NTII) < 0,001$, $p(NTI - NTII) = 0,540$; Lindenhof-Nicht-Kenner (N=241): Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(VT < NTI) < 0,001$, $p(VT < NTII) < 0,001$, $p(NTI - NTII) = 0,660$; ; Lindenhof-Kenner (N=121): Friedmann-Test; $p < 0,001$; Wilcoxon-Test; $p(VT < NTI) < 0,001$, $p(VT < NTII) < 0,001$, $p(NTI - NTII) = 0,655$). Obwohl es scheinbar zwischen den Teilgruppen keine Unterschiede gibt, so zeigen sich leichte Differenzen bezüglich des Geschlechtervergleichs. Die Lindenhof-Kenner zeigen jeweils keine Unterschiede (Mann-Whitney-U-Test, Geschlechtsspezifische Lernunterschiede: Vortest: $p = 0,993$, Nachtest I: $p = 0,743$, Nachtest II: $p = 0,552$). Die Lindenhof-Nicht-Kenner dagegen schon: Mädchen haben im Nachtest II etwas bessere Ergebnisse (Mann-Whitney-U-Test, Geschlechtsspezifische Lernunterschiede: Vortest: $p = 0,176$, Nachtest I: $p = 0,762$, Nachtest II: $p = 0,036$). Eine Berechnung von Treatmentunterschieden erübrigt sich wegen der identischen Behandlung der Schüler im Kindermuseum.

Die affektive Einschätzung der Schüler, geäußert in den gebundenen Aufgabenbeantwortungen im Kontext der Stationenbewertung, war im Nachtest I mit 4,80 und im Nachtest II mit 4,64 sehr positiv und damit deutlich höher als die affektive Gesamtbewertung. Bei dem offenen Item „Willst Du zu Deinem Museumsbesuch noch etwas sagen?“ äußerten sich 51 von 366 Schülern positiv über ihren Besuch des Kindermuseums. Die als letzte Aufgabe ausgeführten Zeichnungen des „Lieblings-Gegenstandes“ der Schüler ergab eine Rangfolge, angeführt von „Kletterbaum“ und „Höhle“, beides Elemente, die ausschließlich Teil des Kindermuseums sind. Kurz: Das Kindermuseum verdient anscheinend seinen Namen. Die Schüler gehen besonders gern in diesen Museumsteil.

Zusammenfassung der Ergebnisse zum Kindermuseum:

- In Gesamtgruppe und Teilgruppen gilt: Erwerb, Persistenz und fehlender Verlust von Wissen.
- Der einzige signifikante kognitive Unterschied zwischen den Geschlechtern findet sich bei Lindenhof-Nicht-Kennern im Nachtest II: Mädchen erbringen bessere Leistungen als Jungen.
- Die affektive Bewertung des Kindermuseums ist besonders positiv.

VI. Diskussion

1. Überlegungen zu Theorie und Aussagebereich

Ausgehend von einer gemäßigten Position der Lerntheorie des Konstruktivismus (Reinmann-Rothmeier & Mandl 2001) wird ein Faktor als wesentlich für Lernprozesse identifiziert: die Selbstbestimmung (vgl. z.B. Killermann 2000, S. 24; vgl. Häußler et al. 1998, S. 158 ff.). Das Maß der Selbstbestimmung wird intentional und kontrolliert variiert, indem drei unterschiedliche unterrichtliche Möglichkeiten entwickelt und umgesetzt werden: Das S-Treatment mit einem hohen Maß an Selbstbestimmung, das SF-Treatment mit mittlerer Selbst- und Fremdbestimmung und das F-Treatment mit dem höchsten Grad an Fremdbestimmung. Drei Schülergruppen bearbeiten im selben Museum mit denselben inhaltlichen Vorgaben (identische Dioramen und Schaukästen) unterschiedliche Aufgabentypen.

Die Untersuchung ist im Spannungsfeld der lerntheoretischen Diskussion um Instruktion und Konstruktion zu sehen. Der organisierte Museumsbesuch erfolgt im instruktionalen Rahmen einer Schulveranstaltung. Die Schüler gehen zuerst zur Schule, bevor sie zum Museum fahren. Ihr Lehrer ist während der ganzen Veranstaltung anwesend. Sie werden vom Versuchsleiter und den Betreuerinnen erwartet und es wird ihnen eine bestimmte Organisation für den Museumsbesuch vorgegeben. Des Weiteren – vielleicht besonders entscheidend – wissen Schüler der fünften Jahrgangsstufe in etwa, was man von ihnen bei der Schulveranstaltung „Museumsbesuch“ erwartet. Den Schülern ist klar, dass sie statt eines Schultages im Klassenraum einen Schultag im Museum verbringen. Das ist vermutlich ein entscheidender Punkt. Bei der Erwartungshaltung „Wandertag“ seitens der Schüler ist die Perzeption möglicherweise völlig anders und die Ergebnisse der hier evaluierten Prozesse wären es ebenfalls. Die bisher aufgeführten Punkte gelten für alle drei Schülergruppen (S-, SF- und F-Schüler). Erst bei der Bearbeitung der Stationen variieren die drei Treatments.

Während auf der Ebene der Erkenntnistheorie des Konstruktivismus alle Lernprozesse konstruktivistisch sind, demgemäß auch alle Lernprozesse die durch die in dieser Untersuchung umgesetzten drei Treatments vermittelt werden, entspricht keines dieser Treatments einer Reinformen von Lerntheorien des Konstruktivismus.

In dieser Untersuchung wird demnach im instruktionalen Rahmen einer Schulveranstaltung ein Merkmal variiert, das für das Lernen im Sinne von Lerntheorien des Konstruktivismus für wichtig erachtet wird. Die drei Treatments können auf einem „bipolaren Kontinuum“ zwischen Instruktion und Konstruktion angeordnet werden (vgl. Hoops 1998, S. 247; Wilde et al. 2003): das S-Treatment in der Nähe der Konstruktion, das F-Treatment in der Nähe der Instruktion und das SF-Treatment im Bereich der Mitte zwischen Instruktion und Konstruktion.

Sehr grundsätzlich lässt sich vorab fragen, ob empirische Untersuchungen zu im weitesten Sinne konstruktivistischen Theorien, Evaluierungen einer „Philosophie“, per se sinnvoll sein können und ob nicht das hier angewandte Instrumentarium quantitativer Verfahren für konstruktivistische Ansätze an sich ungeeignet ist.

Entscheidend dafür ist es, die in der Einleitung vorgestellten Ebenen zu differenzieren. Zunächst gibt es „Konstruktivismus“ als Erkenntnistheorie. Auf diese Theorie stützen sich Modelle der pädagogischen Psychologie zu Lerntheorien des Konstruktivismus. Sicher kann man dies kritisieren oder an sich in Frage stellen. Die Tatsache bleibt, dass es Lerntheorien gibt, die sich auf den Konstruktivismus berufen, sich als konstruktivistisch bezeichnen und als Lerntheorie des Konstruktivismus Kriterien für Lernprozesse liefern. Diese Kriterien lassen sich didaktisch operationalisieren in Form von unterschiedlichen unterrichtlichen Verfahren. Die Umsetzungen dieser Verfahren kann man in Form von Leistungs- und Persönlichkeitstests evaluieren. Das hier variierte Kriterium ist die Selbststeuerung bezüglich der Lerntheorie der gemäßigten konstruktivistischen Position nach Reinmann-Rothmeier & Mandl (2001). Die mit Hilfe von drei Treatments empirisch ermittelten Wirkungen des Museumsbesuchs lassen grundsätzlich Rückschlüsse auf das Kriterium Selbstbestimmung zu und damit auf die Lerntheorie des gemäßigten Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001). Ob sich diese Rückschlüsse auf die ursprüngliche Theorie, die Erkenntnistheorie des Konstruktivismus, beziehen lassen, ist in der Tat zu bezweifeln (vgl. Einleitung).

2. Die zentralen Anliegen der Untersuchung

Evaluiert werden die Folgen der drei Treatments auf kognitiver und affektiver Ebene. Als Kurzform kann man die Hauptfragen der Untersuchung folgendermaßen formulieren:

Was bringt Selbstbestimmung auf kognitiver Ebene?

- Welches Maß an Selbstbestimmung bewirkt guten kurzfristigen Lernerfolg, welches die besten Behaltensleistungen?

Was bringt Selbstbestimmung auf affektiver Ebene?

- Entwickeln die Schüler durch mehr Selbstbestimmung bezüglich des Lernortes bzw. des Lerngegenstandes eher positive Attribuierungen?

3. Diskussion methodischer Aspekte

Auf methodischer Ebene ist vorab zu diskutieren, ob in den drei Treatments tatsächlich die Selbstbestimmung der Probanden als zentraler Einfluss variiert wird und dementsprechend die unterschiedlichen Ergebnisse auf diese Ursache zurückgeführt werden können. Ist die Untersuchung bezüglich dieses Punktes valide?

Man könnte sich auf die „ultimative Lösung“ des Validitätsproblems zurückziehen: „Ein Merkmal ist, was Merkmal-Tests messen.“ Auf den vorliegenden Fall angewandt, würde diese „Lösung“ lauten: „Selbstbestimmung ist das, was in den drei Treatments variiert wird.“ Diese Selbstbezüglichkeit sagt nichts über das Merkmal an sich aus, nützt darum der Klärung nicht wirklich.

Nach Reinmann-Rothmeier & Mandl (2001, S. 632 f.) gewährt selbstbestimmtes Lernen dem Lerner einen hohen Grad an Autonomie. Einzelne Handlungsabschnitte können fremdbestimmt sein. Entscheidend sind Wahlmöglichkeiten des Lernenden bezüglich des Lernprozesses selbst, der Organisation seiner Einzelschritte und des Ziels. Betrachtet man „die Papierform“ der Untersuchung, die Ebene der Planung von außen, so sind diese Bedingungen formal und inhaltlich beim S-Treatment erfüllt, beim SF-Treatment zu einem gewissen Grade und beim F-Treatment gar nicht.

Wie erste Auswertungen zu Verhaltensbeobachtungen aus Videoanalysen andeuten und weitere Informationen über die Umsetzung aus Eindrücken von Betreuerinnen, begleitenden Lehrern und dem Untersuchungsleiter, entsprechen die Schüler den durch die Treatments an sie gestellten Erwartungen. Folgende Beobachtungen wurden gemacht: S-Schüler arbeiten intensiver zusammen, F-Schüler arbeiten öfter alleine. S-Schüler beginnen verzögert, erst nach Diskussionen, mit der Bearbeitung der Stationen, F-Schüler beginnen sofort. S-Schüler schreiben bis zum Ende der Bearbeitungszeit, F-Schüler sind z. T. eher fertig. SF-Schüler liegen jeweils dazwischen: Sie arbeiten öfter zusammen als F-Schüler, beginnen die Bearbeitung der Stationen nicht verzögert und nutzen meist die gesamte Zeit. Diese Beobachtungen stützen den Eindruck, dass die Schüler die Treatments gemäß der Kriterien von Reinmann-Rothmeier & Mandl (2001; vgl. Einleitung Kapitel 2.3.5.), in erwarteter Art und Weise annehmen.

Wie ein Schüler jedoch seinen individuellen Grad an Selbstbestimmung empfindet, ist anhand dieser Informationen nicht zu entscheiden. Die obigen Beobachtungen beziehen sich auf Begleitumstände, die eine bestimmte Rezeption der Aufgaben wahrscheinlich erscheinen lassen. Man kann nicht ausschließen, dass der instruktionale Rahmen, die neuen Betreuer in ungewohnter Umgebung oder soziale Faktoren in der Gruppe das Selbstbestimmungserleben einzelner Schüler einschränken. Legt man die Definition von Reinmann-Rothmeier & Mandl (2001, S. 633) zugrunde (entscheidend für selbstbestimmtes Lernen sind echte Wahlmöglichkeiten des Lerner), so ist bei den drei Treatments ein unterschiedliches Maß an Selbstbestimmung gegeben.

Inwieweit ein unterschiedliches Maß erlebter Selbstbestimmung realisiert wird, soll im Zusammenhang mit der Selbstbestimmungstheorie von Deci & Ryan (1993) und der Untersuchung von Grolnick & Ryan (1987; vgl. Deci & Ryan 1993) diskutiert werden. Die Untersuchung von Grolnick & Ryan (1987) ähnelt der vorliegenden Untersuchung in einigen Aspekten und soll darum ausführlich analysiert werden. Grolnick und Ryan (1987) untersuchten die Autonomie von Lernern. Drei Lernsituationen wurden geschaffen (vgl. Deci & Ryan 1993, S. 234 f.): eine „nicht direktive“ (Gruppe 1), in der Schüler einen Text lesen und berichten, wie interessant sie diesen finden, eine „direktive“, „autonomieunterstützende“ (Gruppe 2), in der Schüler einen Text lesen und

Lehrer ihnen signalisieren, dass sie persönlich an ihrem Lernfortschritt interessiert sind, und eine „direktive“, „kontrollierende“ (Gruppe 3), in der Schüler einen Text lesen und Lehrer ihnen vorher „mitteilen, dass sie die Lernergebnisse prüfen und benoten“ werden. Zwei Ebenen kognitiver Leistungen werden gemessen: Reproduktion und konzeptuelles Wissen. Beim kurzfristigen reproduktiven Wissen schneidet Gruppe 1 (nicht direktiv) signifikant schlechter ab als die Gruppen 2 und 3, so wie die S-Gruppe in Nachtest I (*KogMit18*) in der vorliegenden Untersuchung im Vergleich mit SF- und F- Schülern. Die schlechteste Behaltensleistung zeigt Gruppe 3 (kontrollierend), nicht unähnlich der Gruppe F in Nachtest II (*KogMit18*). Das beste Follow-up-Ergebnis zeigt Gruppe 2 (autonomieunterstützend). Im konzeptuellen Wissen schneiden die Probanden der 3. Gruppe (kontrollierend) am schwächsten ab, so wie die F-Schüler im Nachtest I (*KogO*).

Die beiden Untersuchungen, die von Grolnick & Ryan (1987) und die vorliegende Studie, die aus unterschiedlichen Wissenschaftsbereichen - der Psychologie und der Fachdidaktik - stammen, sich auf verschiedene Theorien - die Selbstbestimmungstheorie nach Deci & Ryan (1993) und die gemäßigte Lerntheorie des Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001) - berufen, treffen sich in ihrem Hauptthema, Autonomie bzw. Selbstbestimmung. Grolnick & Ryan (1987) untersuchten jedoch das Autonomieempfinden der Probanden.

Sie ließen die Schüler identische Aufgaben bearbeiten und versetzten sie in unterschiedliche Situationen der Autonomiewahrnehmung. In Gruppe 1 (nicht direktiv) wird den Probanden sehr klar und nachvollziehbar ein hohes Maß an Autonomie signalisiert. Ihre Meinung ist gefragt und sie dürfen erzählen, was sie interessant finden. Gruppe 3 (kontrollierend) wird in eine Situation schulischer Instruktion gebracht, inklusive schulischer Kontrollmechanismen und Leistungsdruck mit expliziter Androhung von Belohnung und Strafe, nämlich durch die Ankündigung der Benotung. Gruppe 2 (autonomieunterstützend) dagegen wird mit persönlicher Anteilnahme (und mehr Freundlichkeit?) behandelt; sicher angenehm für die Probanden und förderlich für den Lernprozess.

Ein direkter Zusammenhang mit echter Autonomie des Lernalters, im Sinne von mehr realen Entscheidungsmöglichkeiten, ist nicht zu sehen. Erinnerung sei in diesem Zusammenhang nochmals an die Definition von Reinmann-Rothmeier & Mandl (2001, S. 633): „Voraussetzung [für selbstbestimmtes Lernen] (...) ist, dass der Lernende Wahlmöglichkeiten hat und ihm die Möglichkeit gegeben wird, selbst etwas zu bewirken.“ Nach dieser Definition unterscheidet sich die Selbstbestimmung der Lerner bei Grolnick & Ryan (1987) der Gruppen 1, 2 und 3 nicht. In der vorliegenden Untersu-

chung wird dagegen tatsächlich – mit all den schon berichteten methodischen Begleiterscheinungen – die Selbstbestimmung der Schüler variiert. Die Schüler werden nicht in psychologische Situationen unterschiedlicher Wahrnehmung von Selbstbestimmung gebracht, sondern sie können real in unterschiedlichem Maße selbst bestimmen. Trotz dieses zentralen Unterschieds sind die gefundenen Parallelen bemerkenswert. Gruppe 1, nicht direktiv behandelt, entspricht der S-Gruppe. Sogar die Aufgabenstellung („Berichte, wie interessant der Text war!“ versus „Erzähle, was Dich interessiert!“) ist ähnlich. Gruppe 3, direktiv und kontrollierend behandelt, entspricht den instruktional behandelten Probanden (F-Schüler). Die kleinschrittigen Vorgaben könnten eine ähnliche subjektive Situation bedingen wie die direktiv-kontrollierende Lernbedingung bei Grolnick & Ryan. Für Gruppe 2 (autonomieunterstützend) findet sich keine Entsprechung. Dennoch zeigen Probanden der Gruppe 2 wie die der SF-Gruppe im Gesamtbild besonders gute kognitive Leistungen. Allein die Tatsache, ein hohes Maß an Autonomie zu empfinden, scheint keinen guten Einfluss auf kurzfristigen Lernerfolg bei Inhalten, die „nur“ Reproduktion erfordern, zu haben. Sehr viel Kontrolle dagegen scheint längerfristiges Behalten und „nicht-träges Wissen“, dem man in etwa die Ebene „konzeptuelles Wissen“ bzw. die Testmethode *KogO* zurechnen könnte, schlecht zu fördern.

Der obige Einwand, die Bedingungen für Gruppe 2 (autonomieunterstützend) würden die reale Selbstbestimmung der Lerner nicht beeinflussen, ist zu relativieren, berücksichtigt man die „Selbstbestimmungstheorie der Motivation“ von Deci & Ryan (1993, vgl. Deci & Ryan 1985, 2000). Die hier relevanten Aspekte lassen sich verkürzt folgendermaßen darstellen: „In dem Ausmaß, in dem eine motivierte Handlung als frei gewählt erlebt wird, gilt sie als selbstbestimmt oder autonom. In dem Ausmaß, in dem sie als aufgezwungen erlebt wird, gilt sie als kontrolliert“ (Deci & Ryan 1993, S. 225). Dabei gehen Deci & Ryan von einem graduellen Übergang von selbstbestimmtem zu kontrolliertem Verhalten aus, analog zur Sicht von Lernprozessen in einem bipolaren Kontinuum zwischen Instruktion und Konstruktion (vgl. Hoops 1998, S. 247; Wilde et al. 2003). „Intrinsisch motivierte Handlungen repräsentieren den Prototyp selbstbestimmten Verhaltens“ (Deci & Ryan 1993, S. 226). Dieses Verhalten ist autotelischer Natur, wird aus einem Eigenbedürfnis heraus ausgeführt, z.B. Spielen; bei dem untersuchten Museumsbesuch am ehesten im Kindermuseum zu beobachten. (Dort verbringen die meisten Schüler nach dem organisierten und so letztlich fremdbestimmten Unterricht im Museum die Wartezeit bis zur Abfahrt.)

Extrinsisch motiviertes Verhalten wird jedoch nicht notwendigerweise als fremdbestimmt erlebt. Es gibt nach Deci & Ryan (1993, S. 227 f.) eine abgestufte Hierarchie des Erlebens von Selbstbestimmung und Mechanismen zur subjektiven Vereinnahmung extrinsischer Motivation hin zu einem höheren Grad erlebter Selbstbestimmung. Diese Mechanismen können dazu führen, dass extrinsisch motiviertes Verhalten als selbstbestimmt erlebt wird.

Insoweit lässt sich diese „Selbstbestimmungstheorie“ sehr gut mit den drei Treatments S, SF und F vereinbaren. Die unterschiedlichen Grade an Selbstbestimmung sind im instruktionalen Rahmen einer Schulveranstaltung eingebettet, wodurch die schulische Kontrolle, das System von Belohnung, Strafe und aufgezwungenen Zielen implizit in die Untersuchung hinein wirkt. All diese Faktoren führen dazu, dass eine Situation eher als fremdbestimmt erlebt wird (Deci & Ryan 1993, S. 230 f.). Gleichzeitig wird dadurch die intrinsische Motiviertheit negativ beeinflusst. Inwieweit trotz dieser Bedingungen bei den Probanden von echter intrinsischer Motivation gesprochen werden kann, ist nicht entscheidbar. Davon unabhängig können die Treatments unterschiedliche Grade „gefühlter“ Selbstbestimmung für die Schüler bedeuten. Ganz entscheidend für den Grad der subjektiv empfundenen Selbstbestimmung (und die intrinsische Motivation) sind Wahlmöglichkeiten, echte Einflussmöglichkeiten. Daneben zählen Deci & Ryan (1993) weitere Möglichkeiten auf, die intrinsische Motivation positiv beeinflussen, wie das Erleben von Kompetenz und Selbstwirksamkeit oder die Äußerung anerkennender Gefühle. Das letztere lässt sich in unterschiedlicher Weise vermitteln, u. a. auch „autonomiefördernd“. Es ist möglich, autonomiefördernd die intrinsische Motivation positiv zu beeinflussen, so dass das Erleben von Selbstbestimmung gesteigert wird. In diesem Kontext kann man Gruppe 2 aus der Untersuchung von Grolnick & Ryan (1987) als autonomieförderndes Treatment begreifen. Der Ansatzpunkt der vorliegenden Studie dagegen ist, die tatsächlichen Wahlmöglichkeiten der Probanden zu verändern, was nach der Theorie von Deci und Ryan ebenfalls den Grad erlebter Selbstbestimmung beeinflusst.

Fazit: Auch gemäß der Selbstbestimmungstheorie von Deci & Ryan (1993) bieten die drei hier untersuchten Treatments ein unterschiedliches Maß an Selbstbestimmung. Das schließt einen unterschiedlichen Grad an „gefühlter“ Selbstbestimmung mit ein. Dafür sprechen u. a. die ähnlichen Befunde von Grolnick & Ryan (1987), die in ihrer Studie ausschließlich die „gefühlte“ Selbstbestimmung variieren. Unter der Bedingung, dass beide Untersuchungen (die von Grolnick & Ryan (1987) und die vorliegende Studie) nicht mit denselben systematischen Fehlern behaftet sind, stützt die Ähnlichkeit der Ergebnisse bei vergleichbarer Theorie Validität und Objektivität der vorliegenden Studie (vgl. Material und Methode und Bortz & Döring, 1995, S. 180 ff.).

Nach Bortz & Döring (1995, S. 463 ff.) überprüfen statistische Signifikanztests, mit welcher Wahrscheinlichkeit ein gefundenes Ergebnis auftreten kann, „wenn die Populationsverhältnisse der Nullhypothese H_0 entsprechen“. „Ist die Wahrscheinlichkeit kleiner als $\alpha\%$ (in dieser Untersuchung als $\alpha=5\%$ festgelegt), bezeichnet [man] das Stichprobenergebnis als signifikant“. H_0 ist zu verwerfen, die Alternativhypothese H_1 zu akzeptieren. Die Irrtumswahrscheinlichkeit, d.h. die Wahrscheinlichkeit, fälschlicherweise H_0 zu verwerfen (= α -Fehler), beträgt folglich 5%. Eine Fehlentscheidung zugunsten von H_0 gegen H_1 bezeichnet man als β -Fehler (Bortz & Döring 1995, S. 467; vgl. Zöfel 2002, S. 65 ff.). Es liegt in der Natur von Wahrscheinlichkeitstests, α - und β -Fehler nicht ausschließen zu können. Bei sinnvoller Übereinstimmung der statistischen Ergebnisse mit gängigen Theorien oder bei mehreren in sich konsistenten Befunden sind entsprechende Fehlentscheidungen weitgehend auszuschließen. Einen Zweifelsfall in dieser Untersuchung für einen möglichen statistischen Irrtum bietet *KogMit18* / Lindenhof-Nicht-Kenner / Analyse der Einzeltreatments: F- und SF-Schüler zeigen sehr deutliche Unterschiede zwischen Nachtest I und II. H_0 lautet: Die Ergebnisse der Nachtests unterscheiden sich nicht in ihrer zentralen Tendenz. Anhand von Signifikanzen von $p<0,001$, somit bei einem α -Wert von unter 0,1%, wird H_0 zugunsten von H_1 , die Ergebnisse unterscheiden sich, verworfen. Interpretiert werden die Ergebnisse als Verlust von Wissen: Vergessen. Die S-Schüler weichen von diesen Ergebnissen ab: $p=0,067$. H_0 bleibt bestehen, H_1 wird nicht akzeptiert. Im Raum steht ein möglicher β -Fehler, ein irrtümliches Festhalten an der Nullhypothese H_0 . Gestützt wird diese Vermutung durch ...

- das oben berichtete Vergessen der beiden anderen Einzeltreatments der Lindenhof-Nicht-Kenner, SF- und F-Schüler, bei *KogMit18*.
- den Nachweis von Vergessen in der Gesamtgruppe bei *KogMit18*.
- das nachgewiesene Vergessen der S-Schüler bei beiden Teilgruppen, Lindenhof-Kennern und Nicht-Kennern, bei *KogO*.

Solche Ergebnisse, bei denen der Verdacht eines statistischen Irrtums besteht, sollte mit größter Vorsicht diskutiert werden.

4. Bedeutung der kognitiven Ergebnissen

Zunächst zu den kognitiven Ergebnissen der Gesamtgruppe von *KogMit18* und *KogO*. Beide Messinstrumente zeigen, die Schüler erwerben Wissen, behalten einen Teil und vergessen einen Teil. Der Wissensunterschied bei *KogMit18* wie auch bei *KogO* zwischen Vortest und den Nachtests ist nicht sehr hoch, vgl. z.B. Abbildung 40 und 45. Entscheidend für große Unterschiede auf dieser Ebene sind Kongruenz zwischen Treatment und Test bzw. möglichst passgenaue Abstimmung der Tests auf die Inhalte der Treatments. Das ist aus Gründen, die der Hauptintention der Untersuchung immanent sind, nicht in idealer Weise zu leisten (vgl. Wilde et al. 2003). Denn wie schon in Material und Methode, Kapitel 4.2.1. erläutert, ist durch das Gewähren eines unterschiedlichen Maßes an Selbstbestimmung die Vorhersagbarkeit der tatsächlich durch die Probanden bearbeiteten Inhalte nicht identisch. In den Stationen sind alle im Test enthaltenen Inhalte zu erfahren. Die Schaufenster bieten jedoch erheblich mehr Informationen, als für fehlerlose Bearbeitung des Tests nötig sind. Allein in den Legenden der Dioramen werden jeweils 14 bis 17 Arten aufgelistet. Über diese Arten werden im Fragebogen z. T. Details wie Schnabelform, Färbung, etc. oder Beziehungen zueinander wie z.B. Größenverhältnisse oder Unterschiede in der Färbung erfragt. Für die Probanden kommen während des Museumsbesuchs weitere Eindrücke hinzu, nämlich durch Pflanzen, Pilze, andere Strukturelemente und Hintergrundbilder, Erläuterungstexte bei den Schaukästen, weitere – eigentlich zu vernachlässigende – Schaukästen und Dioramen, die nicht zum Thema des Museumsbesuch gemacht wurden, das Terrarium mit den Zwergmäusen, alle Wahrnehmungen aus dem Kindermuseum (Stopfpräparate, Spiele, Kletterbaum, Dachshöhle, etc.), daneben Sinneseindrücke aus dem Außenbereich des Lindenhofs und evtl. neue oder ungewohnte soziale Erfahrungen durch unbekannte Betreuer, möglicherweise ungewohnte Partner in der Kleingruppe oder Wahrnehmungen durch die nicht alltägliche Situation einer Exkursion. Trotz der überschaubaren Auswahl von Schaufenstern und deren nicht sehr komplizierten Strukturierung wirkt in der Intervention „Museumsbesuch“ auf die Schüler eine unüberschaubare Menge an Informationen ein. Die Tests decken nur einen kleinen Anteil der Inhalte aus dem Museumsunterricht ab. Es ist demnach durchaus möglich, dass Schüler etwas lernen, das nicht Teil des Fragebogens ist.

Betrachtet man auf der anderen Seite die Fähigkeiten der Schüler, so werden Effekte unterrichtlicher Lernleistungen oft überschätzt. In einer Studie an Schülern der fünften und sechsten Klasse ($N=460$) zum Thema „Ernährung und Verdauung“ und „Photosynthese“ werden im Mittel 1,0 bis 1,1 Begriffe pro Unterrichtsstunde behalten (Berck 1999, S. 89; vgl. Graf 1989, S. 142 ff.). Im Kontext dieser Daten ist der in dieser Studie erreichte Lernzuwachs akzeptabel, insbesondere da pro Station nur eine Bearbeitungszeit von jeweils zehn Minuten zur Verfügung steht, d.h. im Naturkundemuseum eine reine Arbeitszeit von insgesamt 60 Minuten. Eine längere Bearbeitungsdauer pro Station schien bei F-Schülern zu Langeweile führen zu können, in der Schule ein entscheidender Dissatisfaktor (Begriff adaptiert nach Remer 1994, S. 175 f.), kürzere Dauer schien S-Schüler zu überfordern. Die Zeitspanne von zehn Minuten pro Station hatte sich aufgrund der Vorstudien als günstig erwiesen. Eine Intervention so begrenzten zeitlichen Umfangs kann keine größeren Lernerfolge hervorbringen (vgl. Häußler et al. 1998, S. 149). Zwischen dem „allgemeinen Schulerfolg“ und der „Gesamtdauer des erteilten Unterrichts“ wurde in auf das amerikanische Schulsystem bezogenen Untersuchungen eine positive Korrelation von $r=0,38$ nachgewiesen; in einer europäischen Studie zwischen der Anzahl erteilter Physikwochenstunden während der Schulzeit und den Physikkenntnissen Erwachsener beträgt die Korrelation $r=0,40$ (Häußler 1998, S. 154). Dieser Zusammenhang zwischen Unterrichtszeit und Lernerfolg ist für die hier durchgeführte Untersuchung ebenfalls anzunehmen. Darum ist der oben angesprochene nicht sehr hohe (aber signifikante) Wissensunterschied zwischen Vortest und Nachtests nicht überraschend.

Bei der weiteren Diskussion sind schon auf der Ebene der Gesamtgruppe - von den Teilgruppen sei noch nicht die Rede - die unterschiedlichen Ergebnisse der zwei Messinstrumente (*KogMit18* und *KogO*) miteinander in Einklang zu bringen. *KogMit18* ermittelt das SF-Treatment als günstigste Unterrichtsmethode, *KogO* dagegen das S-Treatment.

4.1. Items mit Antwortvorgabe (*KogMit*)

Der Treatmentgruppenvergleich bei *KogMit18* ermittelt für SF-Schüler die besten kognitiven Ergebnisse (vgl. Abb. 41). Gleichzeitig werden SF-Schüler am meisten durch die Itemkonstruktion begünstigt (vgl. Abb. 42 bis 44). Sind die guten Tester-

gebnisse damit als Artefakte entlarvt? Das abweichende Resultat von *KogO* würde dafür sprechen. Die Ergebnisse der Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner könnte man als Argument dagegen sehen. Denn in beiden Gruppen treten in gleicher Weise itemtypenbezogene Effekte auf (vgl. Tabelle 15 und 19), beide Male hat die Treatmentgruppe SF Vorteile, der Treatmentgruppenvergleich fällt jedoch unterschiedlich aus. Nur bei den Nicht-Kennern dominieren die SF-Probanden, bei den Kennern gibt es zwischen den Treatmentgruppen keine Unterschiede. Man kann argumentieren: Trotz itemtypenbezogener Effekte, die immer SF-Schüler bevorteilen, sind SF-Schüler nicht notwendigerweise im Treatmentgruppenvergleich besser als die übrigen Probanden.

Für eine präzisere Analyse ist erforderlich, sich die Ausgangssituation zu vergegenwärtigen: Einer experimentellen Variierung echter Selbstbestimmung und der empirischen Erfassung des Lernerfolgs ist ein Dilemma immanent: Bei der fremdbestimmten Gruppe der F-Schüler kennt man die kleinschrittigen Instruktionen und kann damit mit einiger Wahrscheinlichkeit abschätzen, worauf die Aufmerksamkeit der Schüler gelenkt wird. Das gilt für Schüler des SF-Treatments schon in geringerem Maße. Bei der selbstbestimmten Gruppe sind die Instruktionen so allgemein, dass man nicht abschätzen kann, worauf die Schüler ihre Aufmerksamkeit lenken. Genau darin besteht das Charakteristikum dieses Treatments, dass das Tun der Schüler gerade nicht kleinschrittig fremdbestimmt wird. S-Schüler beschäftigen sich mit dem, was sie selbst interessant, schön oder schützenswert finden.

Das Dilemma besteht darin, dass die Veränderung der Selbstbestimmung notwendigerweise eine Veränderung des Fokus der Schüler bedingt und damit eine entscheidende Bedingung nicht konstant bleibt. Die Maßnahmen, die es erlauben sollen, dennoch sinnvolle Resultate der empirischen Überprüfung zu erhalten, seien kurz skizziert:

Zuerst wurden die Stationen inhaltlich beschränkt. Trotz der oben referierten Fülle, besteht eine Arbeitsstation jeweils aus einem einzigen „Schaufenster“. Kleinere Einheiten sind im Museum des Lindenhofs aus organisatorischen Gründen nicht zu schaffen. Diese Begrenzung ist unabdingbar, um sinnvoll identische Fragebögen einsetzen zu können. Folgendes Beispiel verdeutlicht dies: Schickt man einen Schü-

ler mit der Aufgabe „Schau Dir an, was Dich interessiert!“ in das Deutsche Museum, so ist es völlig sinnlos, ihn danach zu Details eines bestimmten Exponats zu befragen. Aufgrund der Fülle wird sich der Schüler an das betreffende Exponat kaum erinnern können. Bei vorheriger Beschränkung auf eine kleine Einheit, in der dieses Exponat sich befindet, macht es wesentlich wahrscheinlicher, dass der Schüler sich damit befasst haben wird. In der vorliegenden Untersuchung wurde den Schülern in allen Treatments, auch beim S-Treatment, genau vorgegeben, welche Station sie wie lange bearbeiten mussten. (Eine Station bestand aus einem Schaukasten bzw. Diorama mit entsprechendem Aufgabenpaket. Die Bearbeitungszeit betrug zehn Minuten.)

Die **zweite Maßnahme** bestand in der treatmentbezogenen Itemkonstruktion (ausführliche Erläuterungen vgl. Material und Methoden 4.2.1.). Nachdem auch aufgrund der Begrenztheit der zu bearbeitenden Einheiten alle Inhalte zumindest als implizites Lernziel berücksichtigt sind (vgl. Unterrichtliche Umsetzung 2.3.4. und 2.3.5.2.), werden die Items des Fragebogens auf die Treatments bezogen. Die Grundidee besteht darin, drei Gruppen von Items zu schaffen, sechs F-Items, sechs SF-Items und sechs S-Items, die jeweils einer Treatmentgruppe Vorteile bringen, so dass im gesamten Fragebogenteil KogMit18 alle drei Treatmentgruppen in gleichem Maße bevorteilt bzw. benachteiligt sind.

Damit ist das sich aus dem Dilemma ergebende Problem ein Stück weit gelöst. Dennoch muss man sich darüber im Klaren sein, dass die Ergebnisse in den drei Treatmentgruppen auf unterschiedliche Ursachen zurückgeführt werden können, nämlich Effekte, verursacht durch unterschiedliche Lernwirksamkeit der Treatments („echte Lerneffekte“), und Effekte, die durch die Itemkonstruktion bedingt sein können (Artefakte).

Daraus ergibt sich die Notwendigkeit, einer möglichst beider Ursachen für gemessene „Lerneffekte“ gerecht werdender Interpretation der erhaltenen Ergebnisse. Man kann dies als **dritte Maßnahme** zur Begegnung des Eingangsdilemmas auffassen. Diese Interpretation setzt an folgendem Punkt an: Es sollte möglich sein, aufgrund der Treatments und der Itemkonstruktion begründet vorherzusagen, wie die Ergebnisse bei gleicher Treatmentwirksamkeit aussehen müssten (Modell eines „idealen

Artefakts“). Abweichungen von diesem Szenario sollten auf die unterschiedliche Wirksamkeit der Treatments zurückzuführen sein.

Zur Entwicklung des Modells eines „idealen Artefakts“ sind Szenarien mit hypothetischen Schülern durchzuspielen. Vereinfachend wird die Itemschwierigkeit als identisch angenommen.

Erstes Szenario: Man geht von der Eingangsprämisse aus, Schüler - hypothetische Schüler! - tun genau das, was sie sollen, interessieren sich für alles, lernen alles und vergessen nichts. Das Fragebogenergebnis dieser Bedingungen ist Abbildung 57:

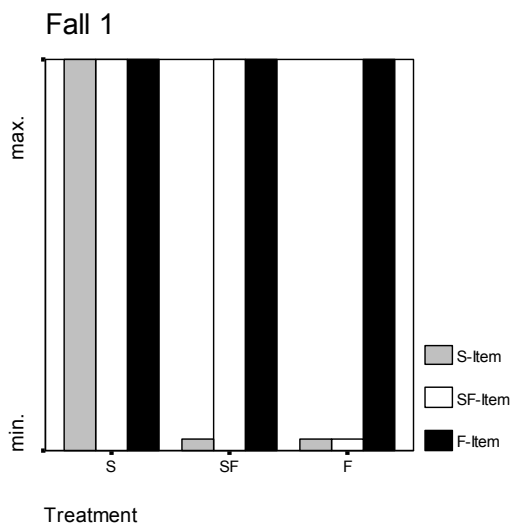


Abbildung 57: Lernerfolg hypothetischer Schüler im 1. Szenario.

S-Schüler können nach dem Treatment alles, denn sie interessieren sich für alles, lernen darum alles und behalten alles perfekt. **SF-Schüler** bearbeiten nicht die Inhalte, die über ihre Aufgaben hinaus gehen, beherrschen darum nicht die Inhalte der S-Items. Für die **F-Treatmentgruppe** gilt dies in analoger Weise: Auch sie beherrschen nur die Inhalte ihres Treatments F, diese jedoch zu 100% (vgl. Abbildung 57).

Zweites Szenario: Justiert man das System an realen Schülern, so muss man die Bedingung, Schüler würden sich für alles interessieren und nichts vergessen, aufgeben. Die Prämissen für die Schüler lauten nun: Schüler tun genau das, was sie sollen, interessieren sich nicht für alles, lernen nicht alles und vergessen. Ergebnis dieser Bedingungen ist die folgende Abbildung 58:

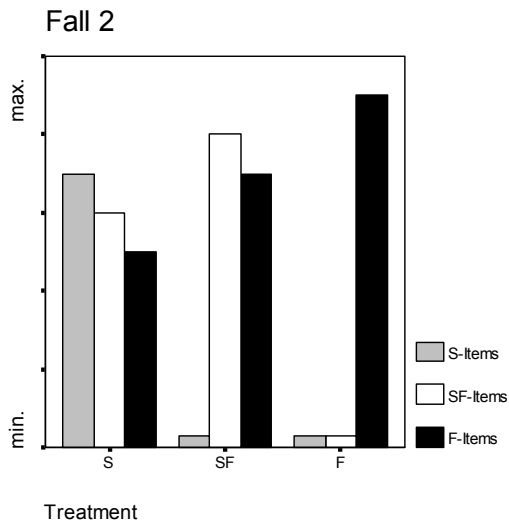


Abbildung 58: Lernerfolg hypothetischer Schüler im 2. Szenario.

F-Schüler tun, was sie sollen, haben darum keine Kenntnisse bei S- und SF-Items. F-Items beantworten sie recht gut, wenngleich durch fehlendes Interesse, schlechtes Lernen oder Behalten der Wert von „100% Lernleistung“ verschieden ist. **SF-Schüler** können SF-Items am besten beantworten, ihre Resultate sollten jedoch etwas schlechter sein, als die der F-Schüler bei den F-Items, denn das SF-Treatment umfasst mehr Inhalte als das F-Treatment. Ein nicht-perfektes Lernen der Schüler bedingt folglich relativ schlechtere Ergebnisse. Der Fokus der SF-Schüler bei F-Items ist vom Kerninhalt ihrer Aufgaben (aufgrund sehr spezieller Inhalte der F-Items) weiter entfernt, aber in den Treatmentaufgaben der SF-Schüler eingeschlossen, so dass das Ergebnis zu F-Items der SF-Schüler relativ zum Ergebnis zu SF-Items etwas schlechter sein sollte. Für die Bearbeiter des **S-Treatments** gilt dies analog: Sie haben die besten Resultate bei S-Items, welche jedoch aufgrund der Stofffülle des S-Treatments relativ schlechter sind als die SF-Items bei SF-Schülern und F-Items bei F-Schülern. S-Schüler können nach wie vor alle Items beantworten, jedoch mit folgender Einschränkung: je spezieller die Items sind, desto schlechter ist die Beantwortung (vgl. Abbildung 58).

Drittes Szenario: Das System ist weiter an realen Schülern auszurichten. Schüler tun meistens nicht genau das, was sie sollen. Die Prämissen für die Schüler lauten dementsprechend: Die Schüler tun nicht genau das, was sie sollen, interessieren sich nicht für alles, lernen nicht alles und vergessen. Das Resultat sieht wie folgt aus (Abbildung 59):

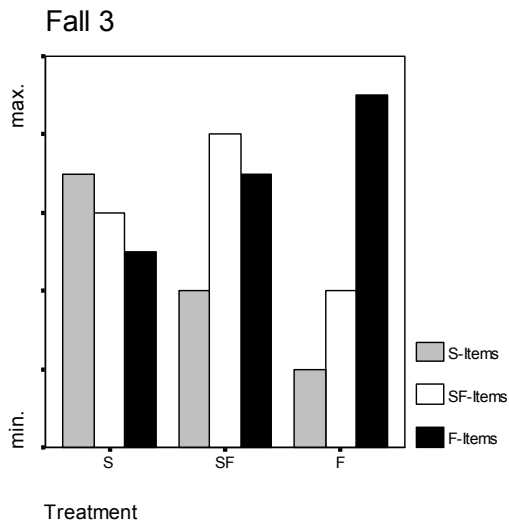


Abbildung 59: Lernerfolg hypothetischer Schüler im 3. Szenario.

Für die **S-Schüler** ändert sich gegenüber dem zweiten Szenario nichts. Bearbeiter des **SF-Treatments** beantworten SF-Items am besten, F-Items, die in ihrer Aufgabe eingeschlossen sind, relativ gut und S-Items mäßig, je nach dem Grad ihres über ihre Aufgabe hinaus gehenden Engagements. Für **F-Schüler** gilt: F-Items beherrschen sie am besten. SF-Items lösen sie besser als S-Items, da dieses Treatment mit ihren Aufgaben verwandt ist (vgl. Material und Methoden 4.2.1.; vgl. Abbildung 59).

Vergleicht man nun die gefundenen Ergebnisse in Nachtest I und II mit diesen Szenarien, so sollte man durch Artefakte verursachte Effekte besser einschätzen können. Dieses Modell eines „idealen Artefakts“ suggeriert in den graphischen Darstellungen der drei Szenarien quantitative Aussagen. Dies ist allenfalls auf der Ebene von Relationen zueinander, zwischen Treatmentgruppen und zwischen Lösungen von Itemtypen, zu sehen, keinesfalls als absolute Aussagen. Das Anliegen ist, hypothetische Vorstellungen zu visualisieren, und nicht scheinquantitative Aussagen zu machen. Die Interpretation erfolgt allein auf qualitativer Ebene, ohne sich notwendigerweise bei jeder Aussage auf statistisch abgesicherte Angaben oder echte Modellrechnungen zu beziehen.

Das Modell eines „idealen Artefakts“ ist für alle Ergebnisse von *KogMit18* gültig. Besonders lohnend ist sein Einsatz bei Lindenhof-Nicht-Kennern. Würde man es auf die Gesamtgruppe beziehen, so schlosse man die Lindenhof-Kenner ein, bei denen keine signifikanten Treatmentunterschiede gefunden werden. In der Teilgruppe der

Kenner überlagert anscheinend ein anderer Faktor, die Bekanntheit des Lindenhof, die hier zu diskutierenden treatmentbezogenen Effekte. Ein Einbeziehen dieser Gruppe bedeutete ein Verwässern möglicher Befunde. Die Interpretation erfolgt darum ausschließlich bezüglich der Nicht-Kenner.

Um Vortestunterschiede zwischen den Treatments, die ohnehin nicht signifikant sind, möglichst auszuschließen, werden die Treatmentunterschiede anhand der Wissenszuwächse (z.B. WZW = Nachtest I – Vortest) betrachtet.

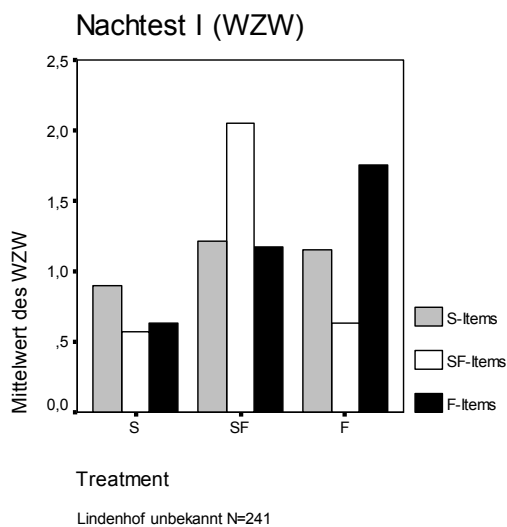


Abbildung 60: Kognitiver Lernerfolg der Treatmentgruppen der Lindenhof-Nicht-Kenner: Wissenszuwachs (WZW = Nachtest I-Ergebnis minus Vortest-Ergebnis) im Nachtest I.

Zum Nachtest I (Anwendung des Modells eines „idealen Artefakts“ auf die Ergebnisse; vgl. Abbildung 60):

1. Im dritten Szenario werden die Schüler am realistischsten eingeschätzt: Schüler lernen und behalten nicht perfekt und sie machen nicht genau das, was man ihnen vorgibt.
2. Wie erwartet, gibt es in den jeweiligen Treatmentgruppen relative Experten für ihre Items: F-Schüler können besonders gut F-Items lösen, SF-Schüler SF-Items und S-Schüler S-Items.
3. Kleinere Abweichungen zum Szenario finden sich bei der Reihung der übrigen Ergebnisse innerhalb der Treatmentgruppen: S-Schüler lösen F-Items etwas besser als SF-Items; SF-Schüler S-Items besser als F-Items und F-Schüler lösen deutlich besser S-Items als SF-Items. Das spricht für unterschiedliche Schwierig-

keit der Items. S-Items sind anscheinend eher leicht, SF-Items relativ am schwersten zu lösen; F-Items liegen in der Mitte.

4. Ganz entscheidende Abweichungen gibt es zu der Höhe der erwarteten Expertenergebnisse. Zwar stimmt das Ergebnis mit dem Szenario überein, dass S-Schüler S-Items schlechter lösen, als F-Schüler F-Items, jedoch die SF-Schüler fallen deutlich aus dem Rahmen: Sie sind in ihrem relativen Expertenergebnis deutlich zu gut und dies trotz der schwierigsten Items.

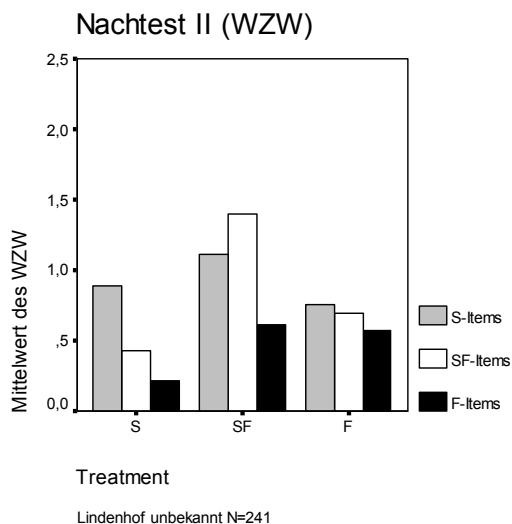


Abbildung 61: Kognitiver Lernerfolg der Treatmentgruppen der Lindenholz-Nicht-Kenner: Wissenszuwachs (WZW = Nachtest II-Ergebnis minus Vortest-Ergebnis) im Nachtest II.

Zum Nachtest II (Anwendung des Modells eines „idealen Artefakts“ auf die Ergebnisse; vgl. Abbildung 61):

1. Wieder kommt nur das dritte Szenario in Frage.
2. Relative Experten gibt es bei S- und bei SF-Schülern, nicht bei den Probanden des F-Treatments. F-Schüler sind bei F-Items am schwächsten, ein unerwartetes Ergebnis.
3. Die Betrachtung der Reihung der übrigen Ergebnisse ergibt bei S-Schülern Übereinstimmung mit dem 3. Szenario des Modells eines „idealen Artefakts“. Bei SF-Schülern sind S- und F-Items im Vergleich zum erwarteten Bild wie schon im Nachtest I vertauscht. F-Schüler zeigen bezüglich der Treatments ein recht einheitliches (niedriges) Niveau, entsprechen dem 3. Szenario gar nicht.

4. Die SF-Schüler übertreffen die übrigen Probanden im Nachtest II noch klarer als im Nachtest I. Sie sind bei allen Itemtypen besser als die übrigen Treatmentgruppen.

Fazit zur Anwendung des Modells eines „idealen Artefakts“ auf die Daten / Ergebnisse der Lindenhof-Nicht-Kenner:

Am ehesten ist das niedrige Niveau der **S-Schüler** durch treatmentbezogene Itemkonstruktion zu erklären. Sie entsprechen dem hypothetischen Szenario in fast idealer Weise. Für diese Treatmentgruppe kann der Fokus der Bearbeitung der Schüler am schlechtesten vorhergesagt werden. Die S-Items werden auf allgemeine Inhalte der „Schaufenster“ bezogen, wobei die Inhalte von SF- und F-Treatment vermieden werden. Es kann darum nicht verwundern, dass S-Items relativ einfacher sind als die Items zu den beiden übrigen Treatments. Das entspricht den Analysen der kognitiven Anforderungsstufen der S-Items nach dem Deutschen Bildungsrat (1970) und der Bloom'schen Taxonomie (Anderson et al. 2001; vgl. Material und Methoden Kapitel 4.2.2.). Diese relativ schwache Lernleistung entspricht zwar der Hypothese H₂ (unterdurchschnittliche kognitive Leistung von S-Schülern), aufgrund der recht guten Übereinstimmung mit dem Modell eines „idealen Artefakts“, bleibt ein gewisses Maß an Unsicherheit bestehen.

Die Ergebnisse der **F-Schüler** im Nachtest I entsprechen in zentralen Punkten den Erwartungen. Der deutliche Abfall im Nachtest II kann nicht auf die Itemkonstruktion zurückgeführt werden. Das F-Treatment bewirkt offensichtlich keine guten Behaltensleistungen. Auffällig ist die Parallele zu Grolnick & Ryan (1987): Die direktiv-kontrollierend behandelte Gruppe zeigte ebenfalls schlechte Ergebnisse im Follow-up-Test (siehe oben). Die Hypothese H₁ (bessere Leistungen bei *KogMit* als bei *KogO*) ist noch nicht bewertbar.

Das insgesamt hohe Niveau der **SF-Schüler** in Nachtest I und II muss sicher z. T. auf die Itemkonstruktion zurückgeführt werden. Die herausragende Position im Nachtest II - die SF-Schüler haben bei SF-Items mit weitem Abstand die besten Ergebnisse und bei den übrigen Itemtypen sind sie besser als die jeweiligen „Experten“ - kann nicht durch die Itemkonstruktion allein bedingt sein. SF-Items sind relativ schwerer als S- und F-Items. Das wird gestützt durch die Ergebnisse der Analyse zu kognitiven

Anforderungsstufen der SF-Items (vgl. Material und Methoden Kapitel 4.2.2). Das SF-Treatment führt bei *KogMit18* zu besonders guten Resultaten. Zumindest teilweise ist Hypothese H₃ (beste kognitive Lernergebnisse) damit bestätigt.

Im Kontext dieser Überlegungen sind die Ergebnisse der F-Schüler, insbesondere im Nachtest II, nicht überzeugend. Man kann dieses Treatment nur sehr eingeschränkt empfehlen. Die Ergebnisse der SF-Schüler sind höher als erwartet, v. a. das Ergebnissniveau im Nachtest II überzeugt. Die unbefriedigenden Resultate der S-Probanden erscheinen im Lichte obiger Überlegungen erklärbar. Möglicherweise wäre es lohnend, die S-Probanden nach Leistungsfähigkeit zu gruppieren. Mandl et al. (1992) beschreibt Lernergebnisse von Probanden, die zwei unterschiedliche Lernumgebungen nutzten: „hypertext versus simultaneous graphical text presentation“ (Mandl et al. 1992, S. 70). Probanden „mit besseren generellen Lernvoraussetzungen“ (Mandl et al. 1997, S. 176) haben in der Hypertext-Lernumgebung bessere Resultate als in der Lernumgebung mit simultaner Textdarbietung. Die weniger leistungsfähige Gruppe lernt mit der simultanen Textdarbietung besser. Beim vorliegenden Ansatz kann man analoge Effekte erwarten. Somit ist es wahrscheinlich, dass das vorliegende Ergebnis der S-Gruppe eine Mischbewertung darstellt, von Schülern, die den Anforderungen, des S-Treatments nicht gewachsen sind, und folglich schlechte Resultate haben und von Schülern, die sehr gut mit diesem Treatment zurecht kommen und optimal gefördert werden. Eine statistische Überprüfung zeigt: Es gibt bei der Gruppe der S-Schüler keine zweigipfelige Verteilung der Beantwortung, die auf Könnern und Nicht-Könnern hindeuten würde, denn beide Messinstrumente *KogMit18* und *KogO* weisen keine signifikant von der Normalverteilung abweichende Verteilungen auf (Kolmogorov-Smirnov-Test; *KogMit18*: $p(\text{NT I})=0,124$, $p(\text{NT II})=0,326$; *KogO*: $p(\text{NT I})=0,207$, $p(\text{NT II})=0,484$; $N=122$). Die nach Mandls Ergebnissen formulierte Hypothese, es gäbe „S-Treatment-Könnern“ und „S-Treatment-Nicht-Könnern“, simplifiziert zu stark. Man findet keine klare Trennung in zwei Leistungsgruppen.

Des Weiteren ist zu bedenken, dass Lernen im F- und im SF-Treatment eher dem gewohnten schulischen Arbeiten entspricht. Trotz der vermehrten Umsetzung freier Arbeitsmethoden in der Grundschule „als Antwort auf den Individualisierungsanspruch heutiger Kinder“ (vgl. Schorch 1998, S. 55) - die Probanden sind ja erst vor

einigen Monaten aus der Grundschule entlassen worden - könnten schlechtere Ergebnisse im S-Treatment der fehlenden Vertrautheit mit Selbstbestimmung geschuldet sein. Die TIMS-Studie beklagt zu deutlich lehrerzentrierten Unterricht (Baumert et al. 1997, S. 231 f.). Als Folgestudie könnte man in einer Längsschnittuntersuchung, die den Schülern die Möglichkeit gibt, sich an diese Bedingungen zu gewöhnen, unter Einbeziehung von Metaebenen, die den Schülern erlauben, ihren methodischen Lernfortschritt zu erkennen und zu beobachten (vgl. Reinmann-Rothmeier & Mandl 2001, S. 616 f.), diesen Punkt aufgreifen. Es ist zu erwarten, dass sich bei größerer Vertrautheit mit freieren Lernmethoden, die guten Lernergebnisse weiter in Richtung S-Treatment verschieben.

Gräsel (1997, S. 96, vgl. Fischer et al. 1997) sieht v. a. den Gegensatz zwischen systematischer und problemorientierter Wissensvermittlung. Problemorientierte Verfahren leiten sich aus Instruktionsmodellen des Situierten Lernens ab und sollen zu anwendbarem Wissen führen, verwendet als Gegensatz zum „trägen Wissen“ (Reinmann-Rothmeier & Mandl 1994, S. 43), das bei traditionellen Methoden der Wissensvermittlung als besonderes Problem identifiziert wurde. Nach Gräsel liegt die Hauptschwierigkeit der problemorientierten Lerner in ihrer Überforderung in authentischen Situationen. Das könnte erklären, warum S-Schüler, denen der Lindenhof unbekannt war, im Treatmentgruppenvergleich relativ schlechter abschneiden als S-Schüler, denen der Lindenhof bekannt war, die keine Treatmentunterschiede bei *KogMit18* zeigen. Die Lindenhof-Kenner könnten schon in höherem Maße über Strategien sinnvoller Bearbeitung verfügen und damit in Situation größerer Selbstbestimmung weniger überfordert sein.

Einen anderen Ansatzpunkt bietet die „cognitive load theory“ (CLT; Owen & Sweller 1985; Sweller et al. 1990; Carlson et al. 2003). Im Zentrum der Theorie steht die kognitive Belastung des Arbeitsgedächtnisses. Zu kleinschrittige und umfangreiche Instruktionen, z.B. bei der Lösung mathematischer Aufgaben (Owen & Sweller 1985), führen nach der CLT zu so großer Belastung des Arbeitsgedächtnisses, dass diese eigentlich zur Unterstützung des Lernalters gedachten Zusatzinformationen dem Lernprozess eher schaden. Für die Bearbeitung des eigentlich zu lösenden Problems bleibt zu wenig Kapazität des Arbeitsspeichers übrig. Ähnlich wirken Aufgaben, die zu viele unterschiedliche Informationen bieten, die gleichzeitig zur Lösung - hier von

Aufgaben aus der Chemie - verarbeitet werden müssen (Carlson et al. 2003). Die Aufgaben der drei Treatments, im Lichte der CLT interpretiert, ergeben für das SF-Treatment die geringste Belastung des Arbeitsspeichers. Das F-Treatment bietet Aufgaben, die sehr viel Instruktion enthalten, deswegen den Arbeitsspeicher, so wie die vielen Hilfen bei Sweller et al. (1985), wahrscheinlich mehr als nötig belasten. Die Anleitungen beim S-Treatment sind zwar sehr zurückhaltend, belasten den Arbeitsspeicher entsprechend wenig, bieten jedoch noch keine konkreten Ansatzpunkte für die Lösung von Problemen. Die Probanden müssen sich die konkreten Aufgaben erst selbst stellen. Man kann annehmen, dass dies das Maß kognitiver Belastung erhöht. Die relativ besten Ergebnisse der SF-Probanden in beiden Nachtests (*KogMit18*) kann man daher mit der CLT begründen.

4.2. Items ohne Antwortvorgabe (*KogO*)

Erweitert man die Perspektive um das zweite Messinstrument *KogO*, das man als **vierte Maßnahme** zur Reaktion auf das obige Dilemma (mit Veränderung der Selbstbestimmung unvermeidliche Veränderung des Fokus' der Schüler) betrachten kann, so werden Überlegungen, die die Ergebnisse des S-Treatments aufwerten, gestützt. Dieses Messinstrument gibt durch seine sehr offenen Items den Schülern Raum für die Präsentation ihres Wissens. Jeder Schüler kann zeigen, was er weiß. Auf der anderen Seite liegt genau darin eine Schwäche von *KogO*, denn wieder stehen die Schüler vor dem Problem der Selbstbestimmung bzw. der Freiheit, schreiben zu können, was sie für wichtig halten. Damit könnten nun die S-Schüler im Vorteil sein, da sie während des Treatments ähnlich arbeiten. Die relativ besten kognitiven Resultate der S-Probanden bei *KogO* kann man im Kontext dieses Vorteils sehen. Dieser Eindruck verschärft sich, wenn man die Ergebnisse des zweiten Nachtests berücksichtigt. Es wäre möglich, dass die Vorteile durch ähnliche Arbeitsweise in Treatment und Test nach sieben Wochen verblasen. Im Nachtest II sind keine Vorteile der S-Schüler mehr statistisch nachweisbar. Bezüglich *KogO* kann die Hypothese H_2 nicht aufrechterhalten werden. Die Ergebnisse der S-Schüler sind in beiden Nachtests nicht unterdurchschnittlich. H_1 wird bestätigt, da die Ergebnisse der F-Gruppe für *KogO* nicht überzeugen, bei *KogMit18* dagegen das Resultat aus Nachtest I akzeptabel ist. Hypothese H_3 lässt sich bezüglich des Testteils *KogO* nicht stützen. Die Ergebnisse der SF-Probanden sind hier nicht besser als die der S-Schüler.

Dieses Messinstrument ist mehr als *KogMit18* geeignet, konzeptuelles (vgl. Grolnick & Ryan 1987) und anwendungsorientiertes (=“nicht-träges“; vgl. Gruber et al. 2000) Wissen zu erfassen. Die Art der Quantifizierung bildet wesentliche Aspekte des Schülerwissens ab. Statistische Kenndaten, insbesondere die Reliabilität, stützen dieses Instrument. Wissen, das ein Schüler frei artikulieren kann, wird ihm in komplexen Anwendungssituationen eher nützen, als Wissen, das nur als Reaktion auf detaillierte Fragen (mit Antwortvorgabe) gezeigt werden kann. „Träges Wissen“ genügt, um in ihrer Komplexität reduzierte Aufgaben zu lösen. Gemeint ist ein „gewissermaßen ‚in vitro‘ erworbenes Wissen“, das „wenig anwendungsbezogen, oft abstrakt und [in] systematisierter Form“ vermittelt wurde (Gruber et al. 2000, S. 139). Dieses „träge Wissen“ ist zwar in einem der Art der Vermittlung analogen Kontext nutzbar, wirkt jedoch in „komplexen, alltagsnahen Problemsituationen“ nicht handlungsleitend; „die Wissensanwendung (...) [gelingt] nicht“ (Gruber et al. 2000, S. 139). In *KogO* gezeigtes Wissen kann Verwendbarkeit in Anwendungssituationen nicht mit Sicherheit voraussagen; Wissen, das für freie Aufgaben, wie *KogO*, nicht abzurufen ist, kann in authentischen Kontexten jedoch wahrscheinlich nicht gezeigt werden. Diese Überlegungen werten die Ergebnisse der Schüler des S-Treatments bei *KogO* auf.

Das Fehlen von Treatmentunterschieden im Nachtest II könnte andeuten, dass der Vorteil der S-Schüler nicht anhaltend ist. Die einmalige Intervention von nur zehn Minuten pro „Schaufenster“ könnte nicht ausreichend gewesen sein, die Eindrücke, die die Schüler direkt nach dem Besuch äußern können, über den langen Zeitraum bis zum Follow-up-Test zu konservieren. In diesem Fall sollten weitere Interventionen wiederholenden Charakters dem Treatment folgen. Das Fehlen von Treatmentunterschieden könnte jedoch auch auf eine Überforderung durch die zu offenen Items im Nachtest II bei *KogO* zurückzuführen sein. Möglicherweise wären bei etwas detaillierteren Items im Nachtest II, die zwar keine Antworten vorgeben, den Schülern jedoch „etwas besser auf die Sprünge helfen“, ausreichend, um im Nachtest II evtl. nach wie vor vorhandene Treatmentunterschiede festzustellen.

4.3. Die Effektstärke von Museumsbesuch und Treatment

Weit mehr als die Hälfte des gemessenen Lernerfolgs ist durch den Museumsbesuch zu erklären ($\omega^2(KogMit18) = 0,57$ (berechnet nach Wolf 2001, S. 99: univariates ω^2 ergibt sich aus dem Prüfwert t des t -Tests, $p < 0,001$) bzw. $\omega^2(KogO) = 0,53$ (analoge Berechnung, $p < 0,001$)), weitere 10% durch Treatment, Geschlecht und Bekanntheit des Lindenhofs¹. Damit ist die Situation „Museumsunterricht“ (gegenübergestellt der Situation „kein Unterricht“) zwar von erheblichem Einfluss für den Lernerfolg, jedoch verglichen mit z.B. Urhahne & Schanze (2003) mit $\omega^2 = 0,80$ nicht sehr hoch. Dies könnte u. a. durch die Themen des Museums bzw. die Inhalte zu erklären sein, die Gegenstand der hier untersuchten Lernprozesse sind. Es handelt sich ausschließlich um „Phänomene heimischer Natur“. Viele der Inhalte haben die Schüler bereits oberflächlich oder vielleicht sogar vertieft kennen gelernt, so dass der Museumsbesuch z. T. eine Erfahrung unter vielen darstellt.

Betrachtet man Einflüsse, die im Rahmen des Museumsbesuches variiert werden, z.B. den Grad der Selbstbestimmung beim Museumsbesuch, so handelt es sich um Faktoren im multikausalen Bedingungsgeflecht von Unterricht. Es kann kaum überraschen, dass sich Selbststeuerung in der Vielzahl der Möglichkeiten und Unwägbarkeiten dieser Situation nicht als ein Effekt von größerer Bedeutung herausstellt. Betrachtet man z.B. $\omega^2(KogMit18) = 0,062$ (berechnet nach Wolf 2001; S. 99: univariates ω^2 ergibt sich aus dem Prüfwert F der ANOVA, $p < 0,001$) im Nachtest I im Kontext von Häußler et al. (1998, S. 152 und 158) - ω^2 liegt hier zwischen 0,02 und 0,03 für „Entdeckendes Lernen“ im Vergleich mit „Normalunterricht“ – so ist die Größe „Treatment“ in dieser Untersuchung zur Erklärung des Lernerfolgs ein beachtenswerter Faktor; nach der Kategorisierung von Häußler et al. (1998, S. 157 f.) in der höchsten Stufe als „Maßnahme mit relativ hoher Effektivität (...)“ einzustufen.

¹ Bestimmt werden diese Effektstärken mittels parametrischer Tests, obwohl nicht in allen Fällen alle Voraussetzungen (Normalverteilung, Varianzhomogenität) gegeben sind (vgl. Material & Methoden).

4.4. Der Einfluss der Bekanntheit des Lindenhofs

Die Kenntnis des Lindenhofs hat für die Ergebnisse entscheidende Bedeutung: Während bei *KogO* zwischen Kennern und Nicht-Kennern die Lernunterschiede marginal sind, differieren die Teilgruppen bezüglich des Fragebogenteils *KogMit18* in fast allen Punkten (vgl. Tabelle 20): Bei Kennern lässt sich im Gegensatz zu Nicht-Kennern kein Vergessen nachweisen, nicht in der Gesamtgruppe und nicht in den Untergruppen der S-, SF- und F-Schüler (vgl. Ergebnisse 1.2.1.1.2.). Ein β -Fehler ist wegen der Konsistenz der Aussagen nicht zu erwarten, wenngleich nicht alle p-Werte Zweifel ausräumen: $p(\text{NT I} - \text{NT II, Gesamtgruppe}) = 0,055$, $p(\text{NT I} - \text{NT II, S-Schüler}) = 0,183$, $p(\text{NT I} - \text{NT II, SF-Schüler}) = 0,342$, $p(\text{NT I} - \text{NT II, F-Schüler}) = 0,329$. Zentral ist das Fehlen statistisch bedeutsamer Unterschiede bei Lindenhof-Kennern zwischen den Treatmentgruppen. Nur bei Nicht-Kennern treten diese Differenzen auf. Interessanterweise sind Effekte treatmentbezogener Itemkonstruktion bei beiden Teilgruppen, Kennern und Nicht-Kennern, in gleichem Maße zu finden. Beide Gruppen lernen und behalten etwa gleichviel.

Es ist demnach offensichtlich nicht so - wie es plausibel wäre -, dass Kenner, deren (empirisch bestätigtes) Charakteristikum unter anderem ist, mehr Vorwissen zu haben, geringeren kurz- oder langfristigen Wissenszuwachs haben, da sie näher am maximal zu erreichenden Niveau sind. Die Hypothese H_4 (Kenner lernen weniger dazu) ist damit zu verwerfen.

Unter Zuhilfenahme der „cognitive load theory“ (CLT; Owen & Sweller 1985; Sweller et al. 1990; Carlson et al. 2003) wird dieses Phänomen erklärlich. Wahrscheinlich beeinflussen die oben beschriebenen Zusatzeindrücke (irrelevante Informationen aus den Schaufenstern und jenseits der Schaufenster) die Nicht-Kenner mehr als die Kenner. Die Kenner wissen eher, welche Zusatzinformationen des Museums einer gedanklichen Beschäftigung lohnen und welche die kognitive Belastung unnötig erhöhen. Beschäftigung mit den „richtigen“ Inhalten erlaubt bessere kognitive Verarbeitung und kann gute Lern- und Behaltensleistungen erklären.

Diese Ergebnisse lassen sich auch im Zusammenhang von Konzepten zur Ambiguitätstoleranz der Schüler diskutieren. Man spricht einer Person geringe Ambiguitätstoleranz (oder Unsicherheitstoleranz) zu, die sich stark von ungewissen Situationen kognitiv beeinflussen lässt (vgl. Reis 1997, S. 8). „Unsicherheitstoleranz nimmt (...) eine Schlüsselstellung im Hinblick auf (...) affektiv relevante Kognitionen ein“ (Reis 1997, S. 8). Damit wären die unterschiedlichen Ergebnisse der Lindenhof-Kenner und der Lindenhof-Nicht-Kenner interpretierbar. Probanden geringer Ambiguitätstoleranz sollte Bekanntheit des Lindenhofs für ihr kognitives Leistungsvermögen am Tag des Besuches nutzen. Die relativ hohen Wissenszuwächse der Lindenhof-Kenner (und Vergessensraten jenseits statistischer Nachweisbarkeit) werden dadurch erklärbar. Diesen Ansatz bestätigt auch Eschenhagen et al. (2001, S. 206): Der Besuch eines außerschulischen Lernortes wirkt am positivsten, „wenn (...) Lernprozesse für Schüler durch einen mittleren Neuigkeitsgrad gekennzeichnet sind“.

Interessant ist das Fehlen von Treatmentunterschieden bei Lindenhof-Kennern. Das könnte bedeuten, dass die Schüler dieser Teilgruppe eine eigene Strategie haben, sich die Inhalte des Museums zu Eigen zu machen. Die unterschiedlichen Treatments, die bei Nicht-Kennern zu deutlich differierenden Resultaten führen, spielen bei den Kennern keine Rolle. In Analogie zur „scientific literacy“ der PISA-Studie (Baumert et al. 2001, S. 195) könnte man eine „museum literacy“ postulieren, eine Eigenschaft bzw. Fähigkeit, die Teilhabe an Kulturgütern bzw. Wissensgütern, die durch Museen bereitgestellt werden, in besonderem Maße erlaubt (vgl. Baumert 2001, S. 195). Lindenhof-Kenner könnten diese „museum literacy“ in höherem Maße besitzen als die Nicht-Kenner und sich darum von instruktionalen Vorgaben, egal ob selbst- oder fremdbestimmt, weniger beeinflussen lassen und stattdessen mehr auf ihre eigene methodische Kompetenz vertrauen, sich einen musealen Inhalt zu erschließen.

Verfolgt man diesen Gedanken weiter, stellt sich die Frage, ob eine vorherige Kenntnis des Lindenhofs tatsächlich für diese Annahme spricht. Bemerkenswert ist, das Fehlen von Zusammenhängen zwischen der Art des Besuches und treatmentbedingter Ergebnisse (vgl. Ergebnisse 1.2.1.1.2.). Wird „museum literacy“ z.B. durch die Teilnahme an einem Kindergeburtstag ausgebildet? Das erscheint zweifelhaft, ist jedoch nicht ganz auszuschließen. Eine weitere Möglichkeit, die den Gedanken der

Existenz einer „museum literacy“ stützen könnte, besteht darin, dass Lindenhof-Kenner v. a. Schüler einer bestimmten gesellschaftlichen Schicht sein könnten, im weiteren Sinne „Kinder des Bildungsbürgertums“. Für die relativ guten Lernleistungen wäre nicht die Kenntnis des Lindenhofs verantwortlich, sondern dieses Merkmal wäre Indikator für eine andere Bedingung („Bildungsbürger“), die entscheidend für den Lernerfolg der Kenner ist. Diese Bedingung kann man umreißen als mehr inhaltliche und methodische Kompetenz der Schüler, aufgrund besserer Förderung durch das Elternhaus. Dies könnte die postulierte „museum literacy“ stützen, hätte jedoch nicht mehr den Charme, diese Kompetenz lustvoll, mühelos und vielleicht durch den einmaligen Besuch eines außerschulischen Lernortes, z.B. bei einem Kindergeburtstag, erwerben zu können.

Ein anderer Erklärungsansatz könnte in unterschiedlicher Interessiertheit der Schüler begründet liegen. Es ist zu postulieren, dass Lindenhof-Kenner das Museum freiwillig aus intrinsischer Motivation besucht haben. In der Zweiteilung in Kenner und Nicht-Kenner könnte die Aufspaltung in an Naturkundemuseen interessierte und nicht-interessierte Schüler verborgen sein. Interesse aber fördert das Lernen.

4.5. Geschlechtsspezifische Phänomene

Bei *KogMit18* sind Jungen besser als Mädchen. Eine Analyse zeigt, dies trifft nur für die Teilgruppe Lindenhof-Nicht-Kenner zu. Bekanntheit lässt diese Differenz verschwinden. Bei *KogO* sind in beiden Teilgruppen die Mädchen besser: Offenere Fragen lassen Mädchen deutlich bessere Ergebnisse als Jungen erzielen. Diese Befunde korrespondieren mit den Ergebnissen zu den Treatmentunterschieden bei *KogO*. Die einzigen statistisch bedeutsamen treatmentbedingten Differenzen lassen sich auf die sehr gut abschneidenden S-Mädchen zurückführen. In gegenläufiger Weise unterscheiden sich die Geschlechter, Hypothese H₅ wird bezüglich der kognitiven Ebene klar bestätigt.

Tendenziell sind geschlechtsspezifische Treatmentunterschiede auch auf affektiver Ebene zu finden. Hier sollte man die Möglichkeit eines statistischen Irrtums nicht ganz ausschließen. Viele Fälle wurden getestet (Gesamtgruppe, sechs Stationen und diese jeweils in beiden Nachtests) und nur relativ wenige signifikante Unterschiede

gefunden. Eine Bonferroni-Korrektur (vgl. Ergebnisse Kapitel 2.2.2.1.) würde die Signifikanzgrenzen so erheblich erhöhen, dass alle Signifikanzen „geopfert“ würden. Darauf wird verzichtet und eine Interpretation der Ergebnisse trotz der Gefahr eines α -Fehlers versucht (vgl. oben): Wieder spielen die Bekanntheit des Lindenhofs und der Grad an Selbstbestimmung eine entscheidende Rolle. In der Gesamtgruppe bewerten F-Jungen mehrfach Stationen signifikant positiver als die übrigen Treatmentgruppen. Bei Mädchen finden sich keine treatmentbedingten Unterschiede. Zieht man die Bekanntheit mit in Betracht, so entsteht aus den statistisch nachweisbaren affektiven Treatmentunterschieden folgendes Bild:

Tabelle 31: Dargestellt sind im Vergleich zu anderen Treatmentgruppen signifikant positive Einschätzungen von Stationen des Museums durch die Treatmentgruppen von Lindenhof-Nicht-Kennern und Lindenhof-Kennern.

	Nicht-Kenner	Kenner
Jungen	F-Schüler	Keine Unterschiede
Mädchen	F-, öfter SF-Schüler	S-Schüler

Nicht-Kenner fühlen sich anscheinend mit einem relativ fremdbestimmtem Treatment wohl: Das ist für Jungen das F-Treatment, für Mädchen eher das SF-Treatment. S-Mädchen, die den Lindenhof vorher kannten, zeigen verglichen mit den übrigen Treatmentgruppen mehrfach die positivsten affektiven Bewertungen. Damit wird Hypothese H₅ zur Existenz von Geschlechtsunterschiede bei der affektiven Wirkung des Museumsbesuchs gestützt.

Folgende Interpretation bietet sich an: Mädchen bringen als S-Schüler gute kognitive Leistungen (KogO, NT I) und bewerten in relativer Selbstbestimmung (als Nicht-Kenner Treatment SF, als Kenner Treatment S) das Museum besonders positiv. „[Positive] emotionale Beteiligung kann das Lernen erheblich verbessern“ (Spitzer 2002, S. 159 f.). Den Jungen in der Untersuchung liegen dagegen eher geschlossene Antwortformate und ein höheres Maß an Fremdbestimmung.

Man kann vermuten, dass Mädchen eine höhere Ambiguitätstoleranz als Jungen besitzen. Sie bewähren sich in Situationen der Unsicherheit, denn sie schneiden in offenen Tests besser ab, und sie fühlen sich in einem Treatment größerer Selbstbestimmung eher wohl als Jungen.

5. Bedeutung der affektiven Dimension

Nach Cooper (1993) ist an konstruktivistischen Leitlinien orientiertes Lernen intrinsisch motiviert. Die Ergebnisse der vorliegenden Untersuchung zeigen keine höheren affektiven Bewertungen der konstruktivistisch orientierten S-Lerner als der übrigen Lerner. Hypothese H₆ ist damit zu verwerfen.

Die Ursachen dafür könnten in einer Überforderung der S-Schüler zu sehen sein. „lower-ability learners perform better in well-structured (...) instructional environments“ (Cooper 1993, S. 13). Neben der messbaren schlechteren Leistung (vgl. Mandl et al. 1992) von „lower-ability learners“ in Lernsituationen, die ihnen komplexe Lernumgebungen bieten, ist gut vorstellbar, dass schwache Schüler im S-Treatment eine subjektive Überforderung wahrnehmen. Ist dies gepaart mit einem Bewusstsein um den instruktionalen schulischen Rahmen und der damit einhergehenden Erwartung kontrollierender Maßnahmen (vgl. Deci & Ryan 1993), wird eine weniger positive affektive Einschätzung durch diese Probanden offensichtlich. Der Effekt könnte durch eine erhöhte Wahrnehmung von Ambiguität (vgl. Reis 1997), verursacht durch die wahrscheinlich ungewohnteren Aufgaben im S-Treatment, verstärkt werden.

Die implizierte Kausalität der Eingangshypothese lässt sich umkehren: Schüler, die positive affektive Bezüge zu Thema und / oder Lernort zeigen und somit als intrinsisch motiviert gelten können, würden sich eher für das S-Treatment entscheiden und mit diesem höchsten Maß an Selbstbestimmung am besten lernen. Weiter gibt es empirische Befunde, die andeuten, dass die Qualität der Stimmung Einfluss auf die Entscheidung, sich interessegeleiteten Aufgaben zu widmen, haben kann (Krapp 1992, S. 40). Das könnte bedeuten, man sollte versuchen, Schüler bezüglich ihrer Motivationslage (intrinsisch motiviert ?) und ihrer Stimmung einzuschätzen und ihnen je nach dieser Einstufung ein unterschiedliches Maß an Selbstbestimmung zu ermöglichen. Dies sollte in idealer Weise unter der Einbeziehung von Metaebenen erfolgen, denn ein Tun, das eigenen Zielvorstellung entspricht, insbesondere wenn diese Zielvorstellungen bewusst sind, kann zu echter Lernfreude führen (vgl. Krapp 1992, S. 39). Nach der Theorie von Csikszentmihalyi (2000) ist bei intrinsisch motivierten Handlungen besonders häufig das Flow-Erlebnis festzustellen, ein besonderer Zustand, in dem eine Person „voll auf eine Aufgabe konzentriert [ist]“, „(...) ganz in der

Handlung [aufgeht] und sich von äußeren und inneren Störfaktoren kaum beeinflussen lässt. Die Zeit ‚vergeht wie im Flug‘ und die Person hat das sichere Gefühl, den Anforderungen der Situation voll gewachsen zu sein und das Geschehen jederzeit im Blick auf die Handlungsziele optimal kontrollieren zu können“ (Krapp 1992, S. 38). In positiven emotionalen Zuständen kann man Lernaufgaben besser lösen. Man tendiert dazu, ungewöhnliche Aspekte in Überlegungen einzubeziehen und die Grenzen dessen, was man bei der Problemstellung für relevant hält, auszuweiten (Krapp 1992, S. 39). Zum Lernen braucht man „innere Beteiligung“, „eine positive Grundstimmung ist (...) gut für das Lernen (Spitzer 2002, S. 157 ff. und S. 164).

Im instruktionalen Rahmen von Schule herrschen demnach keine besonders günstigen Bedingungen für gutes Lernen. „Eine Lernmotivation, die nicht den Prinzipien des individuellen Selbst entspricht, z.B. weil sie von außen aufoktroziert wird, beeinträchtigt die Effektivität des Lernens (...)“ (Deci & Ryan 1993, S. 235). Deci & Ryan (1993, S. 230) berichten weiter, dass vorhandene intrinsische Motivation durch kontrollierend erlebte Bedingungen zerstört werden kann. Dies können offensichtlich negative Einflüsse wie z.B. Strafandrohung und aufgezwungene Ziele sein, aber auch Termindruck, Bewertungen, Auszeichnungen und materielle Belohnung. Das kann beispielsweise bedeuten: Gibt man einer Person eine Belohnung (z.B. Geld) für etwas, das sie ohnehin tun würde, weil sie es gerne tut, wird sie es danach nicht mehr gerne sondern für die Belohnung (z.B. Geld) tun. Das Tun war vorher autotelischer Natur, intrinsisch motiviert und hat sich durch die kontrollierend erlebte Bedingung zu einem Tun aus extrinsischer Motiviertheit gewandelt. Für die Schule bedeutet dies, dass eine ohne diesen instruktionalen Rahmen intrinsisch motivierte Beschäftigung zu einem Tun als Mittel zum Zweck, z.B. um gute Noten zu bekommen, werden kann. Deci & Ryan (1993, S. 235) folgern, „dass benotete Leistungsprüfungen in der Schule, als die am weitesten verbreiteten Mittel zur Kontrolle der Lernmotivation, ‚Schüsse in den Ofen‘ sind. Sie rufen nicht nur negative affektive Reaktionen hervor, sondern bewirken darüber hinaus auch ein qualitativ schlechteres Lernverhalten.“

Ungelöst bleibt das Problem, von Schülern als uninteressant, langweilig oder mühevoll erlebte Inhalte - die aber als intersubjektive Übereinkunft der Gesellschaft, als zu tradierendes Kulturgut festgesetzt werden - „in die Köpfe der Schüler“ zu bringen. Die meisten Schulsysteme begegnen diesem Problem durch instruktionale Rahmen, die kontrollierende Bedingungen wie Leistungstests mit Bewertungen einschließen.

Emotional positive und gleichzeitig kognitiv sehr produktive Zustände wie ein Flow-Erleben ist in der vorliegenden Untersuchung sehr unwahrscheinlich, aufgrund der zeitlich-organisatorischen Vorgaben, durch die z.B. der instruktionale Rahmen deutlich wird, fast auszuschließen. Ziel sollte es dennoch sein, intrinsische Motivation und entsprechende positive emotionale Zustände zuzulassen. Unterricht, der Selbstbestimmung zulässt, könnte Schülern diese Möglichkeit geben.

6. Unterrichtliche Konsequenzen

Die zentrale Frage nach positiver Attribuierung des Museums und möglichst hohem Lernerfolg durch einen Museumsbesuch aufgreifend, kann über unterrichtliche Konsequenzen spekuliert werden. Die Studie fand im Kontext schulischer Bedingungen statt, so dass Vorbehalte aufgrund eines artifiziellen untersuchungsbedingten Rahmens weitgehend auszuschließen sind. Eine Umsetzung dieser auf quantitativen statistischen Methoden beruhenden Aussagen auf das Unterrichten von Individuen sollte behutsam und mit Augenmaß erfolgen.

Der Museumsbesuch war erfolgreich, affektive Wirkungen sehr positiv, (wenngleich vom Treatment fast unabhängig), kognitive Lernfortschritte persistent. Man kann daher den „Museumsunterricht“ selbst als einmalige halbtägige Intervention in einem wenig „spektakulären“ außerschulischen Lernort wie dem Naturkundemuseum des Umweltschutz-Informationszentrums Lindenhof in Bayreuth empfehlen. Die Tatsache, ob man diesen Unterricht durchführt, spielt eine entscheidendere Rolle, als die, wie man ihn gestaltet. Bezogen auf den Faktor der Selbstbestimmung der Schüler, sollte man keine zu deutlichen instruktionalen Bedingungen schaffen. Die Behaltensleistung wird dadurch negativ beeinflusst. Reale Entscheidungsmöglichkeiten, aber auch subjektiv wahrgenommene Autonomie beeinflussen dagegen Lernprozesse positiv. Überraschenderweise lernen Schüler, die den außerschulischen Lernort

schon kannten, nicht schlechter als die, denen der Lindenhof unbekannt war. Die vielfach zu beobachtende Entscheidung von Lehrern, eine durch einen einmaligen Besuch bekannte Einrichtung nicht wiederzubesuchen, könnte demnach grundlegend falsch sein.

Bezüglich der Lernbedingung Selbstbestimmung ist das Ziel von Unterricht entscheidend. Dies gilt sicher nicht nur für Naturkundemuseen sondern auch für den Lernort Schule. Sollen planbar viele Inhalte vermittelt werden, die nach kurzer Zeit im zur Lernsituation ähnlichen Kontext abzuprüfen sind, führen kleinschrittige instruktionale Vorgaben zu recht guten kognitiven Erfolgen. Die schwachen Behaltensleistungen dieses fremdbestimmten Unterrichts bleiben unbemerkt. Diese Situation ist dem vielfach stattfindenden Biologieunterricht mit seiner Stofffülle und den häufigen Prüfungen über Inhalte der letzten Unterrichtsstunde nicht unähnlich. Unterricht, der Lernen höherer kognitiver Anforderungsstufen entspricht, der „(...) weniger Reproduktion (...) dafür aber verstärkt selbst organisiertes Lernen sowie (...) Anwendung von Wissen in den Vordergrund (...)“ stellt (Klautke 2003, S. 33 ff.), sollte sich eher an Prinzipien einer optimalen Balance zwischen Selbst- und Fremdbestimmung bzw. Konstruktion und Instruktion im Sinne des gemäßigten Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001) halten. Ganz entscheidend ist es, interindividuelle Unterschiede der Schüler zu berücksichtigen. Darum ist eine präzise Positionierung auf diesem Kontinuum zwischen Selbst- und Fremdbestimmung bzw. Konstruktion und Instruktion sicher kaum möglich. Das gilt auch, weil abwechselnde Inhalte und ungleiche Unterrichtssituationen in diesem Punkt wahrscheinlich zu unterschiedlichen Resultaten führen.

Die Entscheidung für diese Positionierung liegt nicht nur bei den Lehrpersonen, sondern auch bei den Intentionen des Lehrplans. Neuere Lehrpläne orientieren sich zwar verstärkt daran, didaktische Absichten zu formulieren. Es bleibt jedoch die Tendenz, möglichst viel Wissen zu vermitteln. Eine Alternative könnte sein, sich mehr auf Grundprinzipien des Fachs zu beschränken und größeres Gewicht auf die jeweiligen fachgemäßen Arbeitsweisen und Methoden des Erkenntniserwerbs zu legen. Hierdurch ließen sich vermehrt handlungsorientierte und selbstbestimmte Lernprozesse initiieren und vielleicht effektiver als bisher neues Wissen und Erkennen in vorhandene subjektive Wissens- und Denkstrukturen des Lernenden einbetten. Nachhaltiges Lernen könnte die Folge sein.

VI. Diskussion

1. Überlegungen zu Theorie und Aussagebereich

Ausgehend von einer gemäßigten Position der Lerntheorie des Konstruktivismus (Reinmann-Rothmeier & Mandl 2001) wird ein Faktor als wesentlich für Lernprozesse identifiziert: die Selbstbestimmung (vgl. z.B. Killermann 2000, S. 24; vgl. Häußler et al. 1998, S. 158 ff.). Das Maß der Selbstbestimmung wird intentional und kontrolliert variiert, indem drei unterschiedliche unterrichtliche Möglichkeiten entwickelt und umgesetzt werden: Das S-Treatment mit einem hohen Maß an Selbstbestimmung, das SF-Treatment mit mittlerer Selbst- und Fremdbestimmung und das F-Treatment mit dem höchsten Grad an Fremdbestimmung. Drei Schülergruppen bearbeiten im selben Museum mit denselben inhaltlichen Vorgaben (identische Dioramen und Schaukästen) unterschiedliche Aufgabentypen.

Die Untersuchung ist im Spannungsfeld der lerntheoretischen Diskussion um Instruktion und Konstruktion zu sehen. Der organisierte Museumsbesuch erfolgt im instruktionalen Rahmen einer Schulveranstaltung. Die Schüler gehen zuerst zur Schule, bevor sie zum Museum fahren. Ihr Lehrer ist während der ganzen Veranstaltung anwesend. Sie werden vom Versuchsleiter und den Betreuerinnen erwartet und es wird ihnen eine bestimmte Organisation für den Museumsbesuch vorgegeben. Des Weiteren – vielleicht besonders entscheidend – wissen Schüler der fünften Jahrgangsstufe in etwa, was man von ihnen bei der Schulveranstaltung „Museumsbesuch“ erwartet. Den Schülern ist klar, dass sie statt eines Schultages im Klassenraum einen Schultag im Museum verbringen. Das ist vermutlich ein entscheidender Punkt. Bei der Erwartungshaltung „Wandertag“ seitens der Schüler ist die Perzeption möglicherweise völlig anders und die Ergebnisse der hier evaluierten Prozesse wären es ebenfalls. Die bisher aufgeführten Punkte gelten für alle drei Schülergruppen (S-, SF- und F-Schüler). Erst bei der Bearbeitung der Stationen variieren die drei Treatments.

Während auf der Ebene der Erkenntnistheorie des Konstruktivismus alle Lernprozesse konstruktivistisch sind, demgemäß auch alle Lernprozesse die durch die in dieser Untersuchung umgesetzten drei Treatments vermittelt werden, entspricht keines dieser Treatments einer Reinformen von Lerntheorien des Konstruktivismus.

In dieser Untersuchung wird demnach im instruktionalen Rahmen einer Schulveranstaltung ein Merkmal variiert, das für das Lernen im Sinne von Lerntheorien des Konstruktivismus für wichtig erachtet wird. Die drei Treatments können auf einem „bipolaren Kontinuum“ zwischen Instruktion und Konstruktion angeordnet werden (vgl. Hoops 1998, S. 247; Wilde et al. 2003): das S-Treatment in der Nähe der Konstruktion, das F-Treatment in der Nähe der Instruktion und das SF-Treatment im Bereich der Mitte zwischen Instruktion und Konstruktion.

Sehr grundsätzlich lässt sich vorab fragen, ob empirische Untersuchungen zu im weitesten Sinne konstruktivistischen Theorien, Evaluierungen einer „Philosophie“, per se sinnvoll sein können und ob nicht das hier angewandte Instrumentarium quantitativer Verfahren für konstruktivistische Ansätze an sich ungeeignet ist.

Entscheidend dafür ist es, die in der Einleitung vorgestellten Ebenen zu differenzieren. Zunächst gibt es „Konstruktivismus“ als Erkenntnistheorie. Auf diese Theorie stützen sich Modelle der pädagogischen Psychologie zu Lerntheorien des Konstruktivismus. Sicher kann man dies kritisieren oder an sich in Frage stellen. Die Tatsache bleibt, dass es Lerntheorien gibt, die sich auf den Konstruktivismus berufen, sich als konstruktivistisch bezeichnen und als Lerntheorie des Konstruktivismus Kriterien für Lernprozesse liefern. Diese Kriterien lassen sich didaktisch operationalisieren in Form von unterschiedlichen unterrichtlichen Verfahren. Die Umsetzungen dieser Verfahren kann man in Form von Leistungs- und Persönlichkeitstests evaluieren. Das hier variierte Kriterium ist die Selbststeuerung bezüglich der Lerntheorie der gemäßigten konstruktivistischen Position nach Reinmann-Rothmeier & Mandl (2001). Die mit Hilfe von drei Treatments empirisch ermittelten Wirkungen des Museumsbesuchs lassen grundsätzlich Rückschlüsse auf das Kriterium Selbstbestimmung zu und damit auf die Lerntheorie des gemäßigten Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001). Ob sich diese Rückschlüsse auf die ursprüngliche Theorie, die Erkenntnistheorie des Konstruktivismus, beziehen lassen, ist in der Tat zu bezweifeln (vgl. Einleitung).

2. Die zentralen Anliegen der Untersuchung

Evaluiert werden die Folgen der drei Treatments auf kognitiver und affektiver Ebene. Als Kurzform kann man die Hauptfragen der Untersuchung folgendermaßen formulieren:

Was bringt Selbstbestimmung auf kognitiver Ebene?

- Welches Maß an Selbstbestimmung bewirkt guten kurzfristigen Lernerfolg, welches die besten Behaltensleistungen?

Was bringt Selbstbestimmung auf affektiver Ebene?

- Entwickeln die Schüler durch mehr Selbstbestimmung bezüglich des Lernortes bzw. des Lerngegenstandes eher positive Attribuierungen?

3. Diskussion methodischer Aspekte

Auf methodischer Ebene ist vorab zu diskutieren, ob in den drei Treatments tatsächlich die Selbstbestimmung der Probanden als zentraler Einfluss variiert wird und dementsprechend die unterschiedlichen Ergebnisse auf diese Ursache zurückgeführt werden können. Ist die Untersuchung bezüglich dieses Punktes valide?

Man könnte sich auf die „ultimative Lösung“ des Validitätsproblems zurückziehen: „Ein Merkmal ist, was Merkmal-Tests messen.“ Auf den vorliegenden Fall angewandt, würde diese „Lösung“ lauten: „Selbstbestimmung ist das, was in den drei Treatments variiert wird.“ Diese Selbstbezüglichkeit sagt nichts über das Merkmal an sich aus, nützt darum der Klärung nicht wirklich.

Nach Reinmann-Rothmeier & Mandl (2001, S. 632 f.) gewährt selbstbestimmtes Lernen dem Lerner einen hohen Grad an Autonomie. Einzelne Handlungsabschnitte können fremdbestimmt sein. Entscheidend sind Wahlmöglichkeiten des Lernenden bezüglich des Lernprozesses selbst, der Organisation seiner Einzelschritte und des Ziels. Betrachtet man „die Papierform“ der Untersuchung, die Ebene der Planung von außen, so sind diese Bedingungen formal und inhaltlich beim S-Treatment erfüllt, beim SF-Treatment zu einem gewissen Grade und beim F-Treatment gar nicht.

Wie erste Auswertungen zu Verhaltensbeobachtungen aus Videoanalysen andeuten und weitere Informationen über die Umsetzung aus Eindrücken von Betreuerinnen, begleitenden Lehrern und dem Untersuchungsleiter, entsprechen die Schüler den durch die Treatments an sie gestellten Erwartungen. Folgende Beobachtungen wurden gemacht: S-Schüler arbeiten intensiver zusammen, F-Schüler arbeiten öfter alleine. S-Schüler beginnen verzögert, erst nach Diskussionen, mit der Bearbeitung der Stationen, F-Schüler beginnen sofort. S-Schüler schreiben bis zum Ende der Bearbeitungszeit, F-Schüler sind z. T. eher fertig. SF-Schüler liegen jeweils dazwischen: Sie arbeiten öfter zusammen als F-Schüler, beginnen die Bearbeitung der Stationen nicht verzögert und nutzen meist die gesamte Zeit. Diese Beobachtungen stützen den Eindruck, dass die Schüler die Treatments gemäß der Kriterien von Reinmann-Rothmeier & Mandl (2001; vgl. Einleitung Kapitel 2.3.5.), in erwarteter Art und Weise annehmen.

Wie ein Schüler jedoch seinen individuellen Grad an Selbstbestimmung empfindet, ist anhand dieser Informationen nicht zu entscheiden. Die obigen Beobachtungen beziehen sich auf Begleitumstände, die eine bestimmte Rezeption der Aufgaben wahrscheinlich erscheinen lassen. Man kann nicht ausschließen, dass der instruktionale Rahmen, die neuen Betreuer in ungewohnter Umgebung oder soziale Faktoren in der Gruppe das Selbstbestimmungserleben einzelner Schüler einschränken. Legt man die Definition von Reinmann-Rothmeier & Mandl (2001, S. 633) zugrunde (entscheidend für selbstbestimmtes Lernen sind echte Wahlmöglichkeiten des Lerner), so ist bei den drei Treatments ein unterschiedliches Maß an Selbstbestimmung gegeben.

Inwieweit ein unterschiedliches Maß erlebter Selbstbestimmung realisiert wird, soll im Zusammenhang mit der Selbstbestimmungstheorie von Deci & Ryan (1993) und der Untersuchung von Grolnick & Ryan (1987; vgl. Deci & Ryan 1993) diskutiert werden. Die Untersuchung von Grolnick & Ryan (1987) ähnelt der vorliegenden Untersuchung in einigen Aspekten und soll darum ausführlich analysiert werden. Grolnick und Ryan (1987) untersuchten die Autonomie von Lernern. Drei Lernsituationen wurden geschaffen (vgl. Deci & Ryan 1993, S. 234 f.): eine „nicht direktive“ (Gruppe 1), in der Schüler einen Text lesen und berichten, wie interessant sie diesen finden, eine „direktive“, „autonomieunterstützende“ (Gruppe 2), in der Schüler einen Text lesen und

Lehrer ihnen signalisieren, dass sie persönlich an ihrem Lernfortschritt interessiert sind, und eine „direktive“, „kontrollierende“ (Gruppe 3), in der Schüler einen Text lesen und Lehrer ihnen vorher „mitteilen, dass sie die Lernergebnisse prüfen und benoten“ werden. Zwei Ebenen kognitiver Leistungen werden gemessen: Reproduktion und konzeptuelles Wissen. Beim kurzfristigen reproduktiven Wissen schneidet Gruppe 1 (nicht direktiv) signifikant schlechter ab als die Gruppen 2 und 3, so wie die S-Gruppe in Nachtest I (*KogMit18*) in der vorliegenden Untersuchung im Vergleich mit SF- und F- Schülern. Die schlechteste Behaltensleistung zeigt Gruppe 3 (kontrollierend), nicht unähnlich der Gruppe F in Nachtest II (*KogMit18*). Das beste Follow-up-Ergebnis zeigt Gruppe 2 (autonomieunterstützend). Im konzeptuellen Wissen schneiden die Probanden der 3. Gruppe (kontrollierend) am schwächsten ab, so wie die F-Schüler im Nachtest I (*KogO*).

Die beiden Untersuchungen, die von Grolnick & Ryan (1987) und die vorliegende Studie, die aus unterschiedlichen Wissenschaftsbereichen - der Psychologie und der Fachdidaktik - stammen, sich auf verschiedene Theorien - die Selbstbestimmungstheorie nach Deci & Ryan (1993) und die gemäßigte Lerntheorie des Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001) - berufen, treffen sich in ihrem Hauptthema, Autonomie bzw. Selbstbestimmung. Grolnick & Ryan (1987) untersuchten jedoch das Autonomieempfinden der Probanden.

Sie ließen die Schüler identische Aufgaben bearbeiten und versetzten sie in unterschiedliche Situationen der Autonomiewahrnehmung. In Gruppe 1 (nicht direktiv) wird den Probanden sehr klar und nachvollziehbar ein hohes Maß an Autonomie signalisiert. Ihre Meinung ist gefragt und sie dürfen erzählen, was sie interessant finden. Gruppe 3 (kontrollierend) wird in eine Situation schulischer Instruktion gebracht, inklusive schulischer Kontrollmechanismen und Leistungsdruck mit expliziter Androhung von Belohnung und Strafe, nämlich durch die Ankündigung der Benotung. Gruppe 2 (autonomieunterstützend) dagegen wird mit persönlicher Anteilnahme (und mehr Freundlichkeit?) behandelt; sicher angenehm für die Probanden und förderlich für den Lernprozess.

Ein direkter Zusammenhang mit echter Autonomie des Lernalters, im Sinne von mehr realen Entscheidungsmöglichkeiten, ist nicht zu sehen. Erinnerung sei in diesem Zusammenhang nochmals an die Definition von Reinmann-Rothmeier & Mandl (2001, S. 633): „Voraussetzung [für selbstbestimmtes Lernen] (...) ist, dass der Lernende Wahlmöglichkeiten hat und ihm die Möglichkeit gegeben wird, selbst etwas zu bewirken.“ Nach dieser Definition unterscheidet sich die Selbstbestimmung der Lerner bei Grolnick & Ryan (1987) der Gruppen 1, 2 und 3 nicht. In der vorliegenden Untersu-

chung wird dagegen tatsächlich – mit all den schon berichteten methodischen Begleiterscheinungen – die Selbstbestimmung der Schüler variiert. Die Schüler werden nicht in psychologische Situationen unterschiedlicher Wahrnehmung von Selbstbestimmung gebracht, sondern sie können real in unterschiedlichem Maße selbst bestimmen. Trotz dieses zentralen Unterschieds sind die gefundenen Parallelen bemerkenswert. Gruppe 1, nicht direktiv behandelt, entspricht der S-Gruppe. Sogar die Aufgabenstellung („Berichte, wie interessant der Text war!“ versus „Erzähle, was Dich interessiert!“) ist ähnlich. Gruppe 3, direktiv und kontrollierend behandelt, entspricht den instruktional behandelten Probanden (F-Schüler). Die kleinschrittigen Vorgaben könnten eine ähnliche subjektive Situation bedingen wie die direktiv-kontrollierende Lernbedingung bei Grolnick & Ryan. Für Gruppe 2 (autonomieunterstützend) findet sich keine Entsprechung. Dennoch zeigen Probanden der Gruppe 2 wie die der SF-Gruppe im Gesamtbild besonders gute kognitive Leistungen. Allein die Tatsache, ein hohes Maß an Autonomie zu empfinden, scheint keinen guten Einfluss auf kurzfristigen Lernerfolg bei Inhalten, die „nur“ Reproduktion erfordern, zu haben. Sehr viel Kontrolle dagegen scheint längerfristiges Behalten und „nicht-träges Wissen“, dem man in etwa die Ebene „konzeptuelles Wissen“ bzw. die Testmethode *KogO* zurechnen könnte, schlecht zu fördern.

Der obige Einwand, die Bedingungen für Gruppe 2 (autonomieunterstützend) würden die reale Selbstbestimmung der Lerner nicht beeinflussen, ist zu relativieren, berücksichtigt man die „Selbstbestimmungstheorie der Motivation“ von Deci & Ryan (1993, vgl. Deci & Ryan 1985, 2000). Die hier relevanten Aspekte lassen sich verkürzt folgendermaßen darstellen: „In dem Ausmaß, in dem eine motivierte Handlung als frei gewählt erlebt wird, gilt sie als selbstbestimmt oder autonom. In dem Ausmaß, in dem sie als aufgezwungen erlebt wird, gilt sie als kontrolliert“ (Deci & Ryan 1993, S. 225). Dabei gehen Deci & Ryan von einem graduellen Übergang von selbstbestimmtem zu kontrolliertem Verhalten aus, analog zur Sicht von Lernprozessen in einem bipolaren Kontinuum zwischen Instruktion und Konstruktion (vgl. Hoops 1998, S. 247; Wilde et al. 2003). „Intrinsisch motivierte Handlungen repräsentieren den Prototyp selbstbestimmten Verhaltens“ (Deci & Ryan 1993, S. 226). Dieses Verhalten ist autotelischer Natur, wird aus einem Eigenbedürfnis heraus ausgeführt, z.B. Spielen; bei dem untersuchten Museumsbesuch am ehesten im Kindermuseum zu beobachten. (Dort verbringen die meisten Schüler nach dem organisierten und so letztlich fremdbestimmten Unterricht im Museum die Wartezeit bis zur Abfahrt.)

Extrinsisch motiviertes Verhalten wird jedoch nicht notwendigerweise als fremdbestimmt erlebt. Es gibt nach Deci & Ryan (1993, S. 227 f.) eine abgestufte Hierarchie des Erlebens von Selbstbestimmung und Mechanismen zur subjektiven Vereinnahmung extrinsischer Motivation hin zu einem höheren Grad erlebter Selbstbestimmung. Diese Mechanismen können dazu führen, dass extrinsisch motiviertes Verhalten als selbstbestimmt erlebt wird.

Insoweit lässt sich diese „Selbstbestimmungstheorie“ sehr gut mit den drei Treatments S, SF und F vereinbaren. Die unterschiedlichen Grade an Selbstbestimmung sind im instruktionalen Rahmen einer Schulveranstaltung eingebettet, wodurch die schulische Kontrolle, das System von Belohnung, Strafe und aufgezwungenen Zielen implizit in die Untersuchung hinein wirkt. All diese Faktoren führen dazu, dass eine Situation eher als fremdbestimmt erlebt wird (Deci & Ryan 1993, S. 230 f.). Gleichzeitig wird dadurch die intrinsische Motiviertheit negativ beeinflusst. Inwieweit trotz dieser Bedingungen bei den Probanden von echter intrinsischer Motivation gesprochen werden kann, ist nicht entscheidbar. Davon unabhängig können die Treatments unterschiedliche Grade „gefühlter“ Selbstbestimmung für die Schüler bedeuten. Ganz entscheidend für den Grad der subjektiv empfundenen Selbstbestimmung (und die intrinsische Motivation) sind Wahlmöglichkeiten, echte Einflussmöglichkeiten. Daneben zählen Deci & Ryan (1993) weitere Möglichkeiten auf, die intrinsische Motivation positiv beeinflussen, wie das Erleben von Kompetenz und Selbstwirksamkeit oder die Äußerung anerkennender Gefühle. Das letztere lässt sich in unterschiedlicher Weise vermitteln, u. a. auch „autonomiefördernd“. Es ist möglich, autonomiefördernd die intrinsische Motivation positiv zu beeinflussen, so dass das Erleben von Selbstbestimmung gesteigert wird. In diesem Kontext kann man Gruppe 2 aus der Untersuchung von Grolnick & Ryan (1987) als autonomieförderndes Treatment begreifen. Der Ansatzpunkt der vorliegenden Studie dagegen ist, die tatsächlichen Wahlmöglichkeiten der Probanden zu verändern, was nach der Theorie von Deci und Ryan ebenfalls den Grad erlebter Selbstbestimmung beeinflusst.

Fazit: Auch gemäß der Selbstbestimmungstheorie von Deci & Ryan (1993) bieten die drei hier untersuchten Treatments ein unterschiedliches Maß an Selbstbestimmung. Das schließt einen unterschiedlichen Grad an „gefühlter“ Selbstbestimmung mit ein. Dafür sprechen u. a. die ähnlichen Befunde von Grolnick & Ryan (1987), die in ihrer Studie ausschließlich die „gefühlte“ Selbstbestimmung variieren. Unter der Bedingung, dass beide Untersuchungen (die von Grolnick & Ryan (1987) und die vorliegende Studie) nicht mit denselben systematischen Fehlern behaftet sind, stützt die Ähnlichkeit der Ergebnisse bei vergleichbarer Theorie Validität und Objektivität der vorliegenden Studie (vgl. Material und Methode und Bortz & Döring, 1995, S. 180 ff.).

Nach Bortz & Döring (1995, S. 463 ff.) überprüfen statistische Signifikanztests, mit welcher Wahrscheinlichkeit ein gefundenes Ergebnis auftreten kann, „wenn die Populationsverhältnisse der Nullhypothese H_0 entsprechen“. „Ist die Wahrscheinlichkeit kleiner als $\alpha\%$ (in dieser Untersuchung als $\alpha=5\%$ festgelegt), bezeichnet [man] das Stichprobenergebnis als signifikant“. H_0 ist zu verwerfen, die Alternativhypothese H_1 zu akzeptieren. Die Irrtumswahrscheinlichkeit, d.h. die Wahrscheinlichkeit, fälschlicherweise H_0 zu verwerfen (= α -Fehler), beträgt folglich 5%. Eine Fehlentscheidung zugunsten von H_0 gegen H_1 bezeichnet man als β -Fehler (Bortz & Döring 1995, S. 467; vgl. Zöfel 2002, S. 65 ff.). Es liegt in der Natur von Wahrscheinlichkeitstests, α - und β -Fehler nicht ausschließen zu können. Bei sinnvoller Übereinstimmung der statistischen Ergebnisse mit gängigen Theorien oder bei mehreren in sich konsistenten Befunden sind entsprechende Fehlentscheidungen weitgehend auszuschließen. Einen Zweifelsfall in dieser Untersuchung für einen möglichen statistischen Irrtum bietet *KogMit18* / Lindenhof-Nicht-Kenner / Analyse der Einzeltreatments: F- und SF-Schüler zeigen sehr deutliche Unterschiede zwischen Nachtest I und II. H_0 lautet: Die Ergebnisse der Nachtests unterscheiden sich nicht in ihrer zentralen Tendenz. Anhand von Signifikanzen von $p<0,001$, somit bei einem α -Wert von unter 0,1%, wird H_0 zugunsten von H_1 , die Ergebnisse unterscheiden sich, verworfen. Interpretiert werden die Ergebnisse als Verlust von Wissen: Vergessen. Die S-Schüler weichen von diesen Ergebnissen ab: $p=0,067$. H_0 bleibt bestehen, H_1 wird nicht akzeptiert. Im Raum steht ein möglicher β -Fehler, ein irrtümliches Festhalten an der Nullhypothese H_0 . Gestützt wird diese Vermutung durch ...

- das oben berichtete Vergessen der beiden anderen Einzeltreatments der Lindenhof-Nicht-Kenner, SF- und F-Schüler, bei *KogMit18*.
- den Nachweis von Vergessen in der Gesamtgruppe bei *KogMit18*.
- das nachgewiesene Vergessen der S-Schüler bei beiden Teilgruppen, Lindenhof-Kennern und Nicht-Kennern, bei *KogO*.

Solche Ergebnisse, bei denen der Verdacht eines statistischen Irrtums besteht, sollte mit größter Vorsicht diskutiert werden.

4. Bedeutung der kognitiven Ergebnissen

Zunächst zu den kognitiven Ergebnissen der Gesamtgruppe von *KogMit18* und *KogO*. Beide Messinstrumente zeigen, die Schüler erwerben Wissen, behalten einen Teil und vergessen einen Teil. Der Wissensunterschied bei *KogMit18* wie auch bei *KogO* zwischen Vortest und den Nachtests ist nicht sehr hoch, vgl. z.B. Abbildung 40 und 45. Entscheidend für große Unterschiede auf dieser Ebene sind Kongruenz zwischen Treatment und Test bzw. möglichst passgenaue Abstimmung der Tests auf die Inhalte der Treatments. Das ist aus Gründen, die der Hauptintention der Untersuchung immanent sind, nicht in idealer Weise zu leisten (vgl. Wilde et al. 2003). Denn wie schon in Material und Methode, Kapitel 4.2.1. erläutert, ist durch das Gewähren eines unterschiedlichen Maßes an Selbstbestimmung die Vorhersagbarkeit der tatsächlich durch die Probanden bearbeiteten Inhalte nicht identisch. In den Stationen sind alle im Test enthaltenen Inhalte zu erfahren. Die Schaufenster bieten jedoch erheblich mehr Informationen, als für fehlerlose Bearbeitung des Tests nötig sind. Allein in den Legenden der Dioramen werden jeweils 14 bis 17 Arten aufgelistet. Über diese Arten werden im Fragebogen z. T. Details wie Schnabelform, Färbung, etc. oder Beziehungen zueinander wie z.B. Größenverhältnisse oder Unterschiede in der Färbung erfragt. Für die Probanden kommen während des Museumsbesuchs weitere Eindrücke hinzu, nämlich durch Pflanzen, Pilze, andere Strukturelemente und Hintergrundbilder, Erläuterungstexte bei den Schaukästen, weitere – eigentlich zu vernachlässigende – Schaukästen und Dioramen, die nicht zum Thema des Museumsbesuch gemacht wurden, das Terrarium mit den Zwergmäusen, alle Wahrnehmungen aus dem Kindermuseum (Stopfpräparate, Spiele, Kletterbaum, Dachshöhle, etc.), daneben Sinneseindrücke aus dem Außenbereich des Lindenhofs und evtl. neue oder ungewohnte soziale Erfahrungen durch unbekannte Betreuer, möglicherweise ungewohnte Partner in der Kleingruppe oder Wahrnehmungen durch die nicht alltägliche Situation einer Exkursion. Trotz der überschaubaren Auswahl von Schaufenstern und deren nicht sehr komplizierten Strukturierung wirkt in der Intervention „Museumsbesuch“ auf die Schüler eine unüberschaubare Menge an Informationen ein. Die Tests decken nur einen kleinen Anteil der Inhalte aus dem Museumsunterricht ab. Es ist demnach durchaus möglich, dass Schüler etwas lernen, das nicht Teil des Fragebogens ist.

Betrachtet man auf der anderen Seite die Fähigkeiten der Schüler, so werden Effekte unterrichtlicher Lernleistungen oft überschätzt. In einer Studie an Schülern der fünften und sechsten Klasse ($N=460$) zum Thema „Ernährung und Verdauung“ und „Photosynthese“ werden im Mittel 1,0 bis 1,1 Begriffe pro Unterrichtsstunde behalten (Berck 1999, S. 89; vgl. Graf 1989, S. 142 ff.). Im Kontext dieser Daten ist der in dieser Studie erreichte Lernzuwachs akzeptabel, insbesondere da pro Station nur eine Bearbeitungszeit von jeweils zehn Minuten zur Verfügung steht, d.h. im Naturkundemuseum eine reine Arbeitszeit von insgesamt 60 Minuten. Eine längere Bearbeitungsdauer pro Station schien bei F-Schülern zu Langeweile führen zu können, in der Schule ein entscheidender Dissatisfaktor (Begriff adaptiert nach Remer 1994, S. 175 f.), kürzere Dauer schien S-Schüler zu überfordern. Die Zeitspanne von zehn Minuten pro Station hatte sich aufgrund der Vorstudien als günstig erwiesen. Eine Intervention so begrenzten zeitlichen Umfangs kann keine größeren Lernerfolge hervorbringen (vgl. Häußler et al. 1998, S. 149). Zwischen dem „allgemeinen Schulerfolg“ und der „Gesamtdauer des erteilten Unterrichts“ wurde in auf das amerikanische Schulsystem bezogenen Untersuchungen eine positive Korrelation von $r=0,38$ nachgewiesen; in einer europäischen Studie zwischen der Anzahl erteilter Physikwochenstunden während der Schulzeit und den Physikkenntnissen Erwachsener beträgt die Korrelation $r=0,40$ (Häußler 1998, S. 154). Dieser Zusammenhang zwischen Unterrichtszeit und Lernerfolg ist für die hier durchgeführte Untersuchung ebenfalls anzunehmen. Darum ist der oben angesprochene nicht sehr hohe (aber signifikante) Wissensunterschied zwischen Vortest und Nachtests nicht überraschend.

Bei der weiteren Diskussion sind schon auf der Ebene der Gesamtgruppe - von den Teilgruppen sei noch nicht die Rede - die unterschiedlichen Ergebnisse der zwei Messinstrumente (*KogMit18* und *KogO*) miteinander in Einklang zu bringen. *KogMit18* ermittelt das SF-Treatment als günstigste Unterrichtsmethode, *KogO* dagegen das S-Treatment.

4.1. Items mit Antwortvorgabe (*KogMit*)

Der Treatmentgruppenvergleich bei *KogMit18* ermittelt für SF-Schüler die besten kognitiven Ergebnisse (vgl. Abb. 41). Gleichzeitig werden SF-Schüler am meisten durch die Itemkonstruktion begünstigt (vgl. Abb. 42 bis 44). Sind die guten Tester-

gebnisse damit als Artefakte entlarvt? Das abweichende Resultat von *KogO* würde dafür sprechen. Die Ergebnisse der Teilgruppen der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner könnte man als Argument dagegen sehen. Denn in beiden Gruppen treten in gleicher Weise itemtypenbezogene Effekte auf (vgl. Tabelle 15 und 19), beide Male hat die Treatmentgruppe SF Vorteile, der Teratmentgruppenvergleich fällt jedoch unterschiedlich aus. Nur bei den Nicht-Kennern dominieren die SF-Probanden, bei den Kennern gibt es zwischen den Treatmentgruppen keine Unterschiede. Man kann argumentieren: Trotz itemtypenbezogener Effekte, die immer SF-Schüler bevorteilen, sind SF-Schüler nicht notwendigerweise im Treatmentgruppenvergleich besser als die übrigen Probanden.

Für eine präzisere Analyse ist erforderlich, sich die Ausgangssituation zu vergegenwärtigen: Einer experimentellen Variierung echter Selbstbestimmung und der empirischen Erfassung des Lernerfolgs ist ein Dilemma immanent: Bei der fremdbestimmten Gruppe der F-Schüler kennt man die kleinschrittigen Instruktionen und kann damit mit einiger Wahrscheinlichkeit abschätzen, worauf die Aufmerksamkeit der Schüler gelenkt wird. Das gilt für Schüler des SF-Treatments schon in geringerem Maße. Bei der selbstbestimmten Gruppe sind die Instruktionen so allgemein, dass man nicht abschätzen kann, worauf die Schüler ihre Aufmerksamkeit lenken. Genau darin besteht das Charakteristikum dieses Treatments, dass das Tun der Schüler gerade nicht kleinschrittig fremdbestimmt wird. S-Schüler beschäftigen sich mit dem, was sie selbst interessant, schön oder schützenswert finden.

Das Dilemma besteht darin, dass die Veränderung der Selbstbestimmung notwendigerweise eine Veränderung des Fokus der Schüler bedingt und damit eine entscheidende Bedingung nicht konstant bleibt. Die Maßnahmen, die es erlauben sollen, dennoch sinnvolle Resultate der empirischen Überprüfung zu erhalten, seien kurz skizziert:

Zuerst wurden die Stationen inhaltlich beschränkt. Trotz der oben referierten Fülle, besteht eine Arbeitsstation jeweils aus einem einzigen „Schaufenster“. Kleinere Einheiten sind im Museum des Lindenhofs aus organisatorischen Gründen nicht zu schaffen. Diese Begrenzung ist unabdingbar, um sinnvoll identische Fragebögen einsetzen zu können. Folgendes Beispiel verdeutlicht dies: Schickt man einen Schü-

ler mit der Aufgabe „Schau Dir an, was Dich interessiert!“ in das Deutsche Museum, so ist es völlig sinnlos, ihn danach zu Details eines bestimmten Exponats zu befragen. Aufgrund der Fülle wird sich der Schüler an das betreffende Exponat kaum erinnern können. Bei vorheriger Beschränkung auf eine kleine Einheit, in der dieses Exponat sich befindet, macht es wesentlich wahrscheinlicher, dass der Schüler sich damit befasst haben wird. In der vorliegenden Untersuchung wurde den Schülern in allen Treatments, auch beim S-Treatment, genau vorgegeben, welche Station sie wie lange bearbeiten mussten. (Eine Station bestand aus einem Schaukasten bzw. Diorama mit entsprechendem Aufgabenpaket. Die Bearbeitungszeit betrug zehn Minuten.)

Die **zweite Maßnahme** bestand in der treatmentbezogenen Itemkonstruktion (ausführliche Erläuterungen vgl. Material und Methoden 4.2.1.). Nachdem auch aufgrund der Begrenztheit der zu bearbeitenden Einheiten alle Inhalte zumindest als implizites Lernziel berücksichtigt sind (vgl. Unterrichtliche Umsetzung 2.3.4. und 2.3.5.2.), werden die Items des Fragebogens auf die Treatments bezogen. Die Grundidee besteht darin, drei Gruppen von Items zu schaffen, sechs F-Items, sechs SF-Items und sechs S-Items, die jeweils einer Treatmentgruppe Vorteile bringen, so dass im gesamten Fragebogenteil KogMit18 alle drei Treatmentgruppen in gleichem Maße bevorteilt bzw. benachteiligt sind.

Damit ist das sich aus dem Dilemma ergebende Problem ein Stück weit gelöst. Dennoch muss man sich darüber im Klaren sein, dass die Ergebnisse in den drei Treatmentgruppen auf unterschiedliche Ursachen zurückgeführt werden können, nämlich Effekte, verursacht durch unterschiedliche Lernwirksamkeit der Treatments („echte Lerneffekte“), und Effekte, die durch die Itemkonstruktion bedingt sein können (Artefakte).

Daraus ergibt sich die Notwendigkeit, einer möglichst beider Ursachen für gemessene „Lerneffekte“ gerecht werdender Interpretation der erhaltenen Ergebnisse. Man kann dies als **dritte Maßnahme** zur Begegnung des Eingangsdilemmas auffassen. Diese Interpretation setzt an folgendem Punkt an: Es sollte möglich sein, aufgrund der Treatments und der Itemkonstruktion begründet vorherzusagen, wie die Ergebnisse bei gleicher Treatmentwirksamkeit aussehen müssten (Modell eines „idealen

Artefakts“). Abweichungen von diesem Szenario sollten auf die unterschiedliche Wirksamkeit der Treatments zurückzuführen sein.

Zur Entwicklung des Modells eines „idealen Artefakts“ sind Szenarien mit hypothetischen Schülern durchzuspielen. Vereinfachend wird die Itemschwierigkeit als identisch angenommen.

Erstes Szenario: Man geht von der Eingangsprämisse aus, Schüler - hypothetische Schüler! - tun genau das, was sie sollen, interessieren sich für alles, lernen alles und vergessen nichts. Das Fragebogenergebnis dieser Bedingungen ist Abbildung 57:

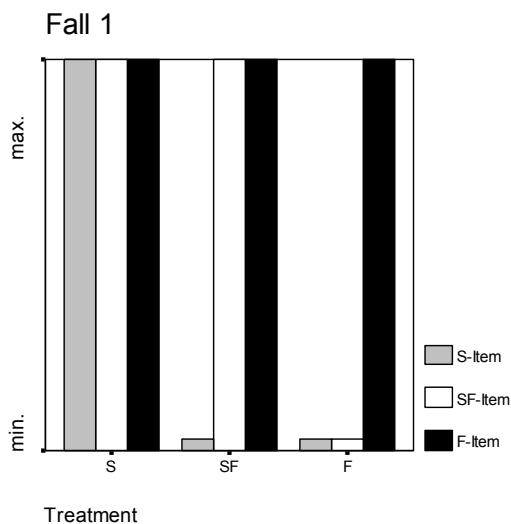


Abbildung 57: Lernerfolg hypothetischer Schüler im 1. Szenario.

S-Schüler können nach dem Treatment alles, denn sie interessieren sich für alles, lernen darum alles und behalten alles perfekt. **SF-Schüler** bearbeiten nicht die Inhalte, die über ihre Aufgaben hinaus gehen, beherrschen darum nicht die Inhalte der S-Items. Für die **F-Treatmentgruppe** gilt dies in analoger Weise: Auch sie beherrschen nur die Inhalte ihres Treatments F, diese jedoch zu 100% (vgl. Abbildung 57).

Zweites Szenario: Justiert man das System an realen Schülern, so muss man die Bedingung, Schüler würden sich für alles interessieren und nichts vergessen, aufgeben. Die Prämissen für die Schüler lauten nun: Schüler tun genau das, was sie sollen, interessieren sich nicht für alles, lernen nicht alles und vergessen. Ergebnis dieser Bedingungen ist die folgende Abbildung 58:

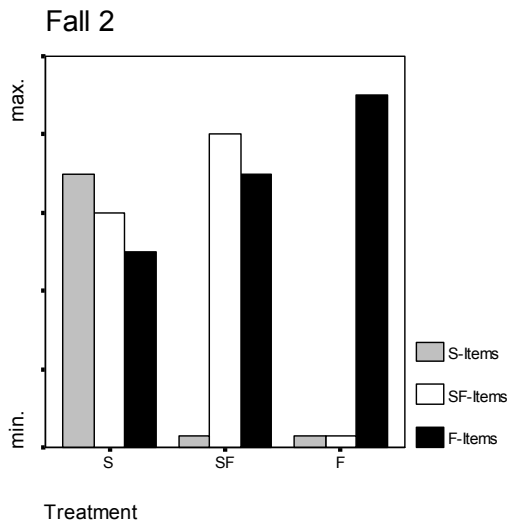


Abbildung 58: Lernerfolg hypothetischer Schüler im 2. Szenario.

F-Schüler tun, was sie sollen, haben darum keine Kenntnisse bei S- und SF-Items. F-Items beantworten sie recht gut, wenngleich durch fehlendes Interesse, schlechtes Lernen oder Behalten der Wert von „100% Lernleistung“ verschieden ist. **SF-Schüler** können SF-Items am besten beantworten, ihre Resultate sollten jedoch etwas schlechter sein, als die der F-Schüler bei den F-Items, denn das SF-Treatment umfasst mehr Inhalte als das F-Treatment. Ein nicht-perfektes Lernen der Schüler bedingt folglich relativ schlechtere Ergebnisse. Der Fokus der SF-Schüler bei F-Items ist vom Kerninhalt ihrer Aufgaben (aufgrund sehr spezieller Inhalte der F-Items) weiter entfernt, aber in den Treatmentaufgaben der SF-Schüler eingeschlossen, so dass das Ergebnis zu F-Items der SF-Schüler relativ zum Ergebnis zu SF-Items etwas schlechter sein sollte. Für die Bearbeiter des **S-Treatments** gilt dies analog: Sie haben die besten Resultate bei S-Items, welche jedoch aufgrund der Stofffülle des S-Treatments relativ schlechter sind als die SF-Items bei SF-Schülern und F-Items bei F-Schülern. S-Schüler können nach wie vor alle Items beantworten, jedoch mit folgender Einschränkung: je spezieller die Items sind, desto schlechter ist die Beantwortung (vgl. Abbildung 58).

Drittes Szenario: Das System ist weiter an realen Schülern auszurichten. Schüler tun meistens nicht genau das, was sie sollen. Die Prämissen für die Schüler lauten dementsprechend: Die Schüler tun nicht genau das, was sie sollen, interessieren sich nicht für alles, lernen nicht alles und vergessen. Das Resultat sieht wie folgt aus (Abbildung 59):

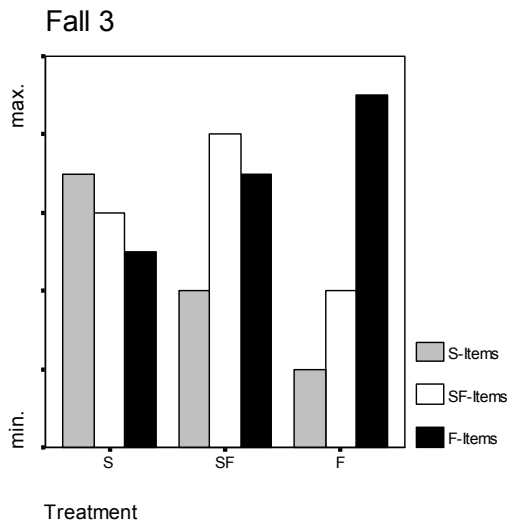


Abbildung 59: Lernerfolg hypothetischer Schüler im 3. Szenario.

Für die **S-Schüler** ändert sich gegenüber dem zweiten Szenario nichts. Bearbeiter des **SF-Treatments** beantworten SF-Items am besten, F-Items, die in ihrer Aufgabe eingeschlossen sind, relativ gut und S-Items mäßig, je nach dem Grad ihres über ihre Aufgabe hinaus gehenden Engagements. Für **F-Schüler** gilt: F-Items beherrschen sie am besten. SF-Items lösen sie besser als S-Items, da dieses Treatment mit ihren Aufgaben verwandt ist (vgl. Material und Methoden 4.2.1.; vgl. Abbildung 59).

Vergleicht man nun die gefundenen Ergebnisse in Nachtest I und II mit diesen Szenarien, so sollte man durch Artefakte verursachte Effekte besser einschätzen können. Dieses Modell eines „idealen Artefakts“ suggeriert in den graphischen Darstellungen der drei Szenarien quantitative Aussagen. Dies ist allenfalls auf der Ebene von Relationen zueinander, zwischen Treatmentgruppen und zwischen Lösungen von Itemtypen, zu sehen, keinesfalls als absolute Aussagen. Das Anliegen ist, hypothetische Vorstellungen zu visualisieren, und nicht scheinquantitative Aussagen zu machen. Die Interpretation erfolgt allein auf qualitativer Ebene, ohne sich notwendigerweise bei jeder Aussage auf statistisch abgesicherte Angaben oder echte Modellrechnungen zu beziehen.

Das Modell eines „idealen Artefakts“ ist für alle Ergebnisse von *KogMit18* gültig. Besonders lohnend ist sein Einsatz bei Lindenhof-Nicht-Kennern. Würde man es auf die Gesamtgruppe beziehen, so schlosse man die Lindenhof-Kenner ein, bei denen keine signifikanten Treatmentunterschiede gefunden werden. In der Teilgruppe der

Kenner überlagert anscheinend ein anderer Faktor, die Bekanntheit des Lindenhof, die hier zu diskutierenden treatmentbezogenen Effekte. Ein Einbeziehen dieser Gruppe bedeutete ein Verwässern möglicher Befunde. Die Interpretation erfolgt darum ausschließlich bezüglich der Nicht-Kenner.

Um Vortestunterschiede zwischen den Treatments, die ohnehin nicht signifikant sind, möglichst auszuschließen, werden die Treatmentunterschiede anhand der Wissenszuwächse (z.B. WZW = Nachtest I – Vortest) betrachtet.

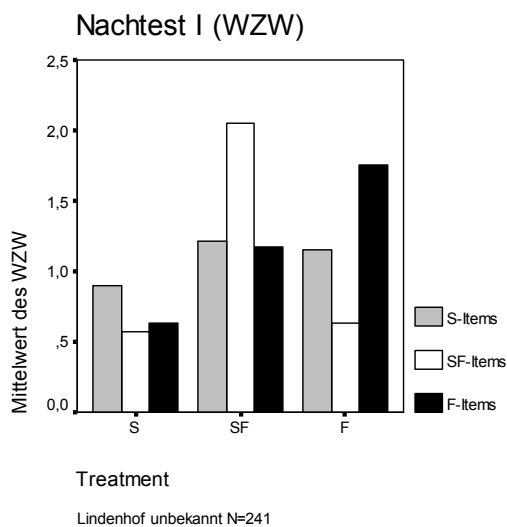


Abbildung 60: Kognitiver Lernerfolg der Treatmentgruppen der Lindenhof-Nicht-Kenner: Wissenszuwachs (WZW = Nachtest I-Ergebnis minus Vortest-Ergebnis) im Nachtest I.

Zum Nachtest I (Anwendung des Modells eines „idealen Artefakts“ auf die Ergebnisse; vgl. Abbildung 60):

1. Im dritten Szenario werden die Schüler am realistischsten eingeschätzt: Schüler lernen und behalten nicht perfekt und sie machen nicht genau das, was man ihnen vorgibt.
2. Wie erwartet, gibt es in den jeweiligen Treatmentgruppen relative Experten für ihre Items: F-Schüler können besonders gut F-Items lösen, SF-Schüler SF-Items und S-Schüler S-Items.
3. Kleinere Abweichungen zum Szenario finden sich bei der Reihung der übrigen Ergebnisse innerhalb der Treatmentgruppen: S-Schüler lösen F-Items etwas besser als SF-Items; SF-Schüler S-Items besser als F-Items und F-Schüler lösen deutlich besser S-Items als SF-Items. Das spricht für unterschiedliche Schwierig-

keit der Items. S-Items sind anscheinend eher leicht, SF-Items relativ am schwersten zu lösen; F-Items liegen in der Mitte.

4. Ganz entscheidende Abweichungen gibt es zu der Höhe der erwarteten Expertenergebnisse. Zwar stimmt das Ergebnis mit dem Szenario überein, dass S-Schüler S-Items schlechter lösen, als F-Schüler F-Items, jedoch die SF-Schüler fallen deutlich aus dem Rahmen: Sie sind in ihrem relativen Expertenergebnis deutlich zu gut und dies trotz der schwierigsten Items.

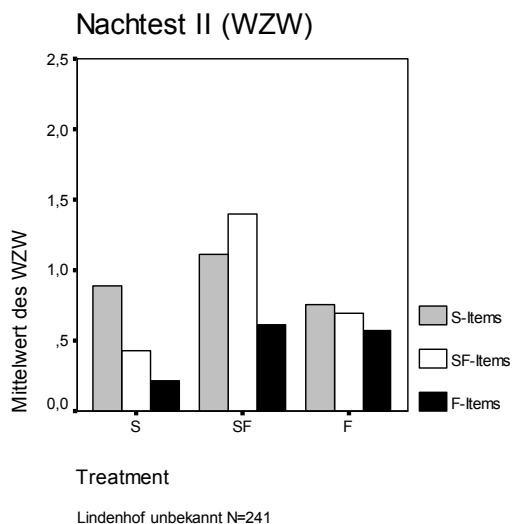


Abbildung 61: Kognitiver Lernerfolg der Treatmentgruppen der Lindenholz-Nicht-Kenner: Wissenszuwachs (WZW = Nachtest II-Ergebnis minus Vortest-Ergebnis) im Nachtest II.

Zum Nachtest II (Anwendung des Modells eines „idealen Artefakts“ auf die Ergebnisse; vgl. Abbildung 61):

1. Wieder kommt nur das dritte Szenario in Frage.
2. Relative Experten gibt es bei S- und bei SF-Schülern, nicht bei den Probanden des F-Treatments. F-Schüler sind bei F-Items am schwächsten, ein unerwartetes Ergebnis.
3. Die Betrachtung der Reihung der übrigen Ergebnisse ergibt bei S-Schülern Übereinstimmung mit dem 3. Szenario des Modells eines „idealen Artefakts“. Bei SF-Schülern sind S- und F-Items im Vergleich zum erwarteten Bild wie schon im Nachtest I vertauscht. F-Schüler zeigen bezüglich der Treatments ein recht einheitliches (niedriges) Niveau, entsprechen dem 3. Szenario gar nicht.

4. Die SF-Schüler übertreffen die übrigen Probanden im Nachtest II noch klarer als im Nachtest I. Sie sind bei allen Itemtypen besser als die übrigen Treatmentgruppen.

Fazit zur Anwendung des Modells eines „idealen Artefakts“ auf die Daten / Ergebnisse der Lindenhof-Nicht-Kenner:

Am ehesten ist das niedrige Niveau der **S-Schüler** durch treatmentbezogene Itemkonstruktion zu erklären. Sie entsprechen dem hypothetischen Szenario in fast idealer Weise. Für diese Treatmentgruppe kann der Fokus der Bearbeitung der Schüler am schlechtesten vorhergesagt werden. Die S-Items werden auf allgemeine Inhalte der „Schaufenster“ bezogen, wobei die Inhalte von SF- und F-Treatment vermieden werden. Es kann darum nicht verwundern, dass S-Items relativ einfacher sind als die Items zu den beiden übrigen Treatments. Das entspricht den Analysen der kognitiven Anforderungsstufen der S-Items nach dem Deutschen Bildungsrat (1970) und der Bloom'schen Taxonomie (Anderson et al. 2001; vgl. Material und Methoden Kapitel 4.2.2.). Diese relativ schwache Lernleistung entspricht zwar der Hypothese H₂ (unterdurchschnittliche kognitive Leistung von S-Schülern), aufgrund der recht guten Übereinstimmung mit dem Modell eines „idealen Artefakts“, bleibt ein gewisses Maß an Unsicherheit bestehen.

Die Ergebnisse der **F-Schüler** im Nachtest I entsprechen in zentralen Punkten den Erwartungen. Der deutliche Abfall im Nachtest II kann nicht auf die Itemkonstruktion zurückgeführt werden. Das F-Treatment bewirkt offensichtlich keine guten Behaltensleistungen. Auffällig ist die Parallele zu Grolnick & Ryan (1987): Die direktiv-kontrollierend behandelte Gruppe zeigte ebenfalls schlechte Ergebnisse im Follow-up-Test (siehe oben). Die Hypothese H₁ (bessere Leistungen bei *KogMit* als bei *KogO*) ist noch nicht bewertbar.

Das insgesamt hohe Niveau der **SF-Schüler** in Nachtest I und II muss sicher z. T. auf die Itemkonstruktion zurückgeführt werden. Die herausragende Position im Nachtest II - die SF-Schüler haben bei SF-Items mit weitem Abstand die besten Ergebnisse und bei den übrigen Itemtypen sind sie besser als die jeweiligen „Experten“ - kann nicht durch die Itemkonstruktion allein bedingt sein. SF-Items sind relativ schwerer als S- und F-Items. Das wird gestützt durch die Ergebnisse der Analyse zu kognitiven

Anforderungsstufen der SF-Items (vgl. Material und Methoden Kapitel 4.2.2). Das SF-Treatment führt bei *KogMit18* zu besonders guten Resultaten. Zumindest teilweise ist Hypothese H₃ (beste kognitive Lernergebnisse) damit bestätigt.

Im Kontext dieser Überlegungen sind die Ergebnisse der F-Schüler, insbesondere im Nachtest II, nicht überzeugend. Man kann dieses Treatment nur sehr eingeschränkt empfehlen. Die Ergebnisse der SF-Schüler sind höher als erwartet, v. a. das Ergebnissniveau im Nachtest II überzeugt. Die unbefriedigenden Resultate der S-Probanden erscheinen im Lichte obiger Überlegungen erklärbar. Möglicherweise wäre es lohnend, die S-Probanden nach Leistungsfähigkeit zu gruppieren. Mandl et al. (1992) beschreibt Lernergebnisse von Probanden, die zwei unterschiedliche Lernumgebungen nutzten: „hypertext versus simultaneous graphical text presentation“ (Mandl et al. 1992, S. 70). Probanden „mit besseren generellen Lernvoraussetzungen“ (Mandl et al. 1997, S. 176) haben in der Hypertext-Lernumgebung bessere Resultate als in der Lernumgebung mit simultaner Textdarbietung. Die weniger leistungsfähige Gruppe lernt mit der simultanen Textdarbietung besser. Beim vorliegenden Ansatz kann man analoge Effekte erwarten. Somit ist es wahrscheinlich, dass das vorliegende Ergebnis der S-Gruppe eine Mischbewertung darstellt, von Schülern, die den Anforderungen, des S-Treatments nicht gewachsen sind, und folglich schlechte Resultate haben und von Schülern, die sehr gut mit diesem Treatment zurecht kommen und optimal gefördert werden. Eine statistische Überprüfung zeigt: Es gibt bei der Gruppe der S-Schüler keine zweigipfelige Verteilung der Beantwortung, die auf Könner und Nicht-Könnern hindeuten würde, denn beide Messinstrumente *KogMit18* und *KogO* weisen keine signifikant von der Normalverteilung abweichende Verteilungen auf (Kolmogorov-Smirnov-Test; *KogMit18*: $p(\text{NT I})=0,124$, $p(\text{NT II})=0,326$; *KogO*: $p(\text{NT I})=0,207$, $p(\text{NT II})=0,484$; $N=122$). Die nach Mandls Ergebnissen formulierte Hypothese, es gäbe „S-Treatment-Könnern“ und „S-Treatment-Nicht-Könnern“, simplifiziert zu stark. Man findet keine klare Trennung in zwei Leistungsgruppen.

Des Weiteren ist zu bedenken, dass Lernen im F- und im SF-Treatment eher dem gewohnten schulischen Arbeiten entspricht. Trotz der vermehrten Umsetzung freierer Arbeitsmethoden in der Grundschule „als Antwort auf den Individualisierungsanspruch heutiger Kinder“ (vgl. Schorch 1998, S. 55) - die Probanden sind ja erst vor

einigen Monaten aus der Grundschule entlassen worden - könnten schlechtere Ergebnisse im S-Treatment der fehlenden Vertrautheit mit Selbstbestimmung geschuldet sein. Die TIMS-Studie beklagt zu deutlich lehrerzentrierten Unterricht (Baumert et al. 1997, S. 231 f.). Als Folgestudie könnte man in einer Längsschnittuntersuchung, die den Schülern die Möglichkeit gibt, sich an diese Bedingungen zu gewöhnen, unter Einbeziehung von Metaebenen, die den Schülern erlauben, ihren methodischen Lernfortschritt zu erkennen und zu beobachten (vgl. Reinmann-Rothmeier & Mandl 2001, S. 616 f.), diesen Punkt aufgreifen. Es ist zu erwarten, dass sich bei größerer Vertrautheit mit freieren Lernmethoden, die guten Lernergebnisse weiter in Richtung S-Treatment verschieben.

Gräsel (1997, S. 96, vgl. Fischer et al. 1997) sieht v. a. den Gegensatz zwischen systematischer und problemorientierter Wissensvermittlung. Problemorientierte Verfahren leiten sich aus Instruktionsmodellen des Situierten Lernens ab und sollen zu anwendbarem Wissen führen, verwendet als Gegensatz zum „trägen Wissen“ (Reinmann-Rothmeier & Mandl 1994, S. 43), das bei traditionellen Methoden der Wissensvermittlung als besonderes Problem identifiziert wurde. Nach Gräsel liegt die Hauptschwierigkeit der problemorientierten Lerner in ihrer Überforderung in authentischen Situationen. Das könnte erklären, warum S-Schüler, denen der Lindenhof unbekannt war, im Treatmentgruppenvergleich relativ schlechter abschneiden als S-Schüler, denen der Lindenhof bekannt war, die keine Treatmentunterschiede bei *KogMit18* zeigen. Die Lindenhof-Kenner könnten schon in höherem Maße über Strategien sinnvoller Bearbeitung verfügen und damit in Situation größerer Selbstbestimmung weniger überfordert sein.

Einen anderen Ansatzpunkt bietet die „cognitive load theory“ (CLT; Owen & Sweller 1985; Sweller et al. 1990; Carlson et al. 2003). Im Zentrum der Theorie steht die kognitive Belastung des Arbeitsgedächtnisses. Zu kleinschrittige und umfangreiche Instruktionen, z.B. bei der Lösung mathematischer Aufgaben (Owen & Sweller 1985), führen nach der CLT zu so großer Belastung des Arbeitsgedächtnisses, dass diese eigentlich zur Unterstützung des Lernalters gedachten Zusatzinformationen dem Lernprozess eher schaden. Für die Bearbeitung des eigentlich zu lösenden Problems bleibt zu wenig Kapazität des Arbeitsspeichers übrig. Ähnlich wirken Aufgaben, die zu viele unterschiedliche Informationen bieten, die gleichzeitig zur Lösung - hier von

Aufgaben aus der Chemie - verarbeitet werden müssen (Carlson et al. 2003). Die Aufgaben der drei Treatments, im Lichte der CLT interpretiert, ergeben für das SF-Treatment die geringste Belastung des Arbeitsspeichers. Das F-Treatment bietet Aufgaben, die sehr viel Instruktion enthalten, deswegen den Arbeitsspeicher, so wie die vielen Hilfen bei Sweller et al. (1985), wahrscheinlich mehr als nötig belasten. Die Anleitungen beim S-Treatment sind zwar sehr zurückhaltend, belasten den Arbeitsspeicher entsprechend wenig, bieten jedoch noch keine konkreten Ansatzpunkte für die Lösung von Problemen. Die Probanden müssen sich die konkreten Aufgaben erst selbst stellen. Man kann annehmen, dass dies das Maß kognitiver Belastung erhöht. Die relativ besten Ergebnisse der SF-Probanden in beiden Nachtests (*KogMit18*) kann man daher mit der CLT begründen.

4.2. Items ohne Antwortvorgabe (*KogO*)

Erweitert man die Perspektive um das zweite Messinstrument *KogO*, das man als **vierte Maßnahme** zur Reaktion auf das obige Dilemma (mit Veränderung der Selbstbestimmung unvermeidliche Veränderung des Fokus' der Schüler) betrachten kann, so werden Überlegungen, die die Ergebnisse des S-Treatments aufwerten, gestützt. Dieses Messinstrument gibt durch seine sehr offenen Items den Schülern Raum für die Präsentation ihres Wissens. Jeder Schüler kann zeigen, was er weiß. Auf der anderen Seite liegt genau darin eine Schwäche von *KogO*, denn wieder stehen die Schüler vor dem Problem der Selbstbestimmung bzw. der Freiheit, schreiben zu können, was sie für wichtig halten. Damit könnten nun die S-Schüler im Vorteil sein, da sie während des Treatments ähnlich arbeiten. Die relativ besten kognitiven Resultate der S-Probanden bei *KogO* kann man im Kontext dieses Vorteils sehen. Dieser Eindruck verschärft sich, wenn man die Ergebnisse des zweiten Nachtests berücksichtigt. Es wäre möglich, dass die Vorteile durch ähnliche Arbeitsweise in Treatment und Test nach sieben Wochen verblasen. Im Nachtest II sind keine Vorteile der S-Schüler mehr statistisch nachweisbar. Bezüglich *KogO* kann die Hypothese H_2 nicht aufrechterhalten werden. Die Ergebnisse der S-Schüler sind in beiden Nachtests nicht unterdurchschnittlich. H_1 wird bestätigt, da die Ergebnisse der F-Gruppe für *KogO* nicht überzeugen, bei *KogMit18* dagegen das Resultat aus Nachtest I akzeptabel ist. Hypothese H_3 lässt sich bezüglich des Testteils *KogO* nicht stützen. Die Ergebnisse der SF-Probanden sind hier nicht besser als die der S-Schüler.

Dieses Messinstrument ist mehr als *KogMit18* geeignet, konzeptuelles (vgl. Grolnick & Ryan 1987) und anwendungsorientiertes (=„nicht-träges“; vgl. Gruber et al. 2000) Wissen zu erfassen. Die Art der Quantifizierung bildet wesentliche Aspekte des Schülerwissens ab. Statistische Kenndaten, insbesondere die Reliabilität, stützen dieses Instrument. Wissen, das ein Schüler frei artikulieren kann, wird ihm in komplexen Anwendungssituationen eher nützen, als Wissen, das nur als Reaktion auf detaillierte Fragen (mit Antwortvorgabe) gezeigt werden kann. „Träges Wissen“ genügt, um in ihrer Komplexität reduzierte Aufgaben zu lösen. Gemeint ist ein „gewissermaßen ‚in vitro‘ erworbenes Wissen“, das „wenig anwendungsbezogen, oft abstrakt und [in] systematisierter Form“ vermittelt wurde (Gruber et al. 2000, S. 139). Dieses „träge Wissen“ ist zwar in einem der Art der Vermittlung analogen Kontext nutzbar, wirkt jedoch in „komplexen, alltagsnahen Problemsituationen“ nicht handlungsleitend; „die Wissensanwendung (...) [gelingt] nicht“ (Gruber et al. 2000, S. 139). In *KogO* gezeigtes Wissen kann Verwendbarkeit in Anwendungssituationen nicht mit Sicherheit voraussagen; Wissen, das für freie Aufgaben, wie *KogO*, nicht abzurufen ist, kann in authentischen Kontexten jedoch wahrscheinlich nicht gezeigt werden. Diese Überlegungen werten die Ergebnisse der Schüler des S-Treatments bei *KogO* auf.

Das Fehlen von Treatmentunterschieden im Nachtest II könnte andeuten, dass der Vorteil der S-Schüler nicht anhaltend ist. Die einmalige Intervention von nur zehn Minuten pro „Schaufenster“ könnte nicht ausreichend gewesen sein, die Eindrücke, die die Schüler direkt nach dem Besuch äußern können, über den langen Zeitraum bis zum Follow-up-Test zu konservieren. In diesem Fall sollten weitere Interventionen wiederholenden Charakters dem Treatment folgen. Das Fehlen von Treatmentunterschieden könnte jedoch auch auf eine Überforderung durch die zu offenen Items im Nachtest II bei *KogO* zurückzuführen sein. Möglicherweise wären bei etwas detaillierteren Items im Nachtest II, die zwar keine Antworten vorgeben, den Schülern jedoch „etwas besser auf die Sprünge helfen“, ausreichend, um im Nachtest II evtl. nach wie vor vorhandene Treatmentunterschiede festzustellen.

4.3. Die Effektstärke von Museumsbesuch und Treatment

Weit mehr als die Hälfte des gemessenen Lernerfolgs ist durch den Museumsbesuch zu erklären ($\omega^2(KogMit18) = 0,57$ (berechnet nach Wolf 2001, S. 99: univariates ω^2 ergibt sich aus dem Prüfwert t des t -Tests, $p < 0,001$) bzw. $\omega^2(KogO) = 0,53$ (analoge Berechnung, $p < 0,001$)), weitere 10% durch Treatment, Geschlecht und Bekanntheit des Lindenhofs¹. Damit ist die Situation „Museumsunterricht“ (gegenübergestellt der Situation „kein Unterricht“) zwar von erheblichem Einfluss für den Lernerfolg, jedoch verglichen mit z.B. Urhahne & Schanze (2003) mit $\omega^2 = 0,80$ nicht sehr hoch. Dies könnte u. a. durch die Themen des Museums bzw. die Inhalte zu erklären sein, die Gegenstand der hier untersuchten Lernprozesse sind. Es handelt sich ausschließlich um „Phänomene heimischer Natur“. Viele der Inhalte haben die Schüler bereits oberflächlich oder vielleicht sogar vertieft kennen gelernt, so dass der Museumsbesuch z. T. eine Erfahrung unter vielen darstellt.

Betrachtet man Einflüsse, die im Rahmen des Museumsbesuches variiert werden, z.B. den Grad der Selbstbestimmung beim Museumsbesuch, so handelt es sich um Faktoren im multikausalen Bedingungsgeflecht von Unterricht. Es kann kaum überraschen, dass sich Selbststeuerung in der Vielzahl der Möglichkeiten und Unwägbarkeiten dieser Situation nicht als ein Effekt von größerer Bedeutung herausstellt. Betrachtet man z.B. $\omega^2(KogMit18) = 0,062$ (berechnet nach Wolf 2001; S. 99: univariates ω^2 ergibt sich aus dem Prüfwert F der ANOVA, $p < 0,001$) im Nachtest I im Kontext von Häußler et al. (1998, S. 152 und 158) - ω^2 liegt hier zwischen 0,02 und 0,03 für „Entdeckendes Lernen“ im Vergleich mit „Normalunterricht“ – so ist die Größe „Treatment“ in dieser Untersuchung zur Erklärung des Lernerfolgs ein beachtenswerter Faktor; nach der Kategorisierung von Häußler et al. (1998, S. 157 f.) in der höchsten Stufe als „Maßnahme mit relativ hoher Effektivität (...)“ einzustufen.

¹ Bestimmt werden diese Effektstärken mittels parametrischer Tests, obwohl nicht in allen Fällen alle Voraussetzungen (Normalverteilung, Varianzhomogenität) gegeben sind (vgl. Material & Methoden).

4.4. Der Einfluss der Bekanntheit des Lindenhofs

Die Kenntnis des Lindenhofs hat für die Ergebnisse entscheidende Bedeutung: Während bei *KogO* zwischen Kennern und Nicht-Kennern die Lernunterschiede marginal sind, differieren die Teilgruppen bezüglich des Fragebogenteils *KogMit18* in fast allen Punkten (vgl. Tabelle 20): Bei Kennern lässt sich im Gegensatz zu Nicht-Kennern kein Vergessen nachweisen, nicht in der Gesamtgruppe und nicht in den Untergruppen der S-, SF- und F-Schüler (vgl. Ergebnisse 1.2.1.1.2.). Ein β -Fehler ist wegen der Konsistenz der Aussagen nicht zu erwarten, wenngleich nicht alle p-Werte Zweifel ausräumen: $p(\text{NT I} - \text{NT II, Gesamtgruppe}) = 0,055$, $p(\text{NT I} - \text{NT II, S-Schüler}) = 0,183$, $p(\text{NT I} - \text{NT II, SF-Schüler}) = 0,342$, $p(\text{NT I} - \text{NT II, F-Schüler}) = 0,329$. Zentral ist das Fehlen statistisch bedeutsamer Unterschiede bei Lindenhof-Kennern zwischen den Treatmentgruppen. Nur bei Nicht-Kennern treten diese Differenzen auf. Interessanterweise sind Effekte treatmentbezogener Itemkonstruktion bei beiden Teilgruppen, Kennern und Nicht-Kennern, in gleichem Maße zu finden. Beide Gruppen lernen und behalten etwa gleichviel.

Es ist demnach offensichtlich nicht so - wie es plausibel wäre -, dass Kenner, deren (empirisch bestätigtes) Charakteristikum unter anderem ist, mehr Vorwissen zu haben, geringeren kurz- oder langfristigen Wissenszuwachs haben, da sie näher am maximal zu erreichenden Niveau sind. Die Hypothese H_4 (Kenner lernen weniger dazu) ist damit zu verwerfen.

Unter Zuhilfenahme der „cognitive load theory“ (CLT; Owen & Sweller 1985; Sweller et al. 1990; Carlson et al. 2003) wird dieses Phänomen erklärlich. Wahrscheinlich beeinflussen die oben beschriebenen Zusatzeindrücke (irrelevante Informationen aus den Schaufenstern und jenseits der Schaufenster) die Nicht-Kenner mehr als die Kenner. Die Kenner wissen eher, welche Zusatzinformationen des Museums einer gedanklichen Beschäftigung lohnen und welche die kognitive Belastung unnötig erhöhen. Beschäftigung mit den „richtigen“ Inhalten erlaubt bessere kognitive Verarbeitung und kann gute Lern- und Behaltensleistungen erklären.

Diese Ergebnisse lassen sich auch im Zusammenhang von Konzepten zur Ambiguitätstoleranz der Schüler diskutieren. Man spricht einer Person geringe Ambiguitätstoleranz (oder Unsicherheitstoleranz) zu, die sich stark von ungewissen Situationen kognitiv beeinflussen lässt (vgl. Reis 1997, S. 8). „Unsicherheitstoleranz nimmt (...) eine Schlüsselstellung im Hinblick auf (...) affektiv relevante Kognitionen ein“ (Reis 1997, S. 8). Damit wären die unterschiedlichen Ergebnisse der Lindenhof-Kenner und der Lindenhof-Nicht-Kenner interpretierbar. Probanden geringer Ambiguitätstoleranz sollte Bekanntheit des Lindenhofs für ihr kognitives Leistungsvermögen am Tag des Besuches nutzen. Die relativ hohen Wissenszuwächse der Lindenhof-Kenner (und Vergessensraten jenseits statistischer Nachweisbarkeit) werden dadurch erklärbar. Diesen Ansatz bestätigt auch Eschenhagen et al. (2001, S. 206): Der Besuch eines außerschulischen Lernortes wirkt am positivsten, „wenn (...) Lernprozesse für Schüler durch einen mittleren Neuigkeitsgrad gekennzeichnet sind“.

Interessant ist das Fehlen von Treatmentunterschieden bei Lindenhof-Kennern. Das könnte bedeuten, dass die Schüler dieser Teilgruppe eine eigene Strategie haben, sich die Inhalte des Museums zu Eigen zu machen. Die unterschiedlichen Treatments, die bei Nicht-Kennern zu deutlich differierenden Resultaten führen, spielen bei den Kennern keine Rolle. In Analogie zur „scientific literacy“ der PISA-Studie (Baumert et al. 2001, S. 195) könnte man eine „museum literacy“ postulieren, eine Eigenschaft bzw. Fähigkeit, die Teilhabe an Kulturgütern bzw. Wissensgütern, die durch Museen bereitgestellt werden, in besonderem Maße erlaubt (vgl. Baumert 2001, S. 195). Lindenhof-Kenner könnten diese „museum literacy“ in höherem Maße besitzen als die Nicht-Kenner und sich darum von instruktionalen Vorgaben, egal ob selbst- oder fremdbestimmt, weniger beeinflussen lassen und stattdessen mehr auf ihre eigene methodische Kompetenz vertrauen, sich einen musealen Inhalt zu erschließen.

Verfolgt man diesen Gedanken weiter, stellt sich die Frage, ob eine vorherige Kenntnis des Lindenhofs tatsächlich für diese Annahme spricht. Bemerkenswert ist, das Fehlen von Zusammenhängen zwischen der Art des Besuches und treatmentbedingter Ergebnisse (vgl. Ergebnisse 1.2.1.1.2.). Wird „museum literacy“ z.B. durch die Teilnahme an einem Kindergeburtstag ausgebildet? Das erscheint zweifelhaft, ist jedoch nicht ganz auszuschließen. Eine weitere Möglichkeit, die den Gedanken der

Existenz einer „museum literacy“ stützen könnte, besteht darin, dass Lindenhof-Kenner v. a. Schüler einer bestimmten gesellschaftlichen Schicht sein könnten, im weiteren Sinne „Kinder des Bildungsbürgertums“. Für die relativ guten Lernleistungen wäre nicht die Kenntnis des Lindenhofs verantwortlich, sondern dieses Merkmal wäre Indikator für eine andere Bedingung („Bildungsbürger“), die entscheidend für den Lernerfolg der Kenner ist. Diese Bedingung kann man umreißen als mehr inhaltliche und methodische Kompetenz der Schüler, aufgrund besserer Förderung durch das Elternhaus. Dies könnte die postulierte „museum literacy“ stützen, hätte jedoch nicht mehr den Charme, diese Kompetenz lustvoll, mühelos und vielleicht durch den einmaligen Besuch eines außerschulischen Lernortes, z.B. bei einem Kindergeburtstag, erwerben zu können.

Ein anderer Erklärungsansatz könnte in unterschiedlicher Interessiertheit der Schüler begründet liegen. Es ist zu postulieren, dass Lindenhof-Kenner das Museum freiwillig aus intrinsischer Motivation besucht haben. In der Zweiteilung in Kenner und Nicht-Kenner könnte die Aufspaltung in an Naturkundemuseen interessierte und nicht-interessierte Schüler verborgen sein. Interesse aber fördert das Lernen.

4.5. Geschlechtsspezifische Phänomene

Bei *KogMit18* sind Jungen besser als Mädchen. Eine Analyse zeigt, dies trifft nur für die Teilgruppe Lindenhof-Nicht-Kenner zu. Bekanntheit lässt diese Differenz verschwinden. Bei *KogO* sind in beiden Teilgruppen die Mädchen besser: Offenere Fragen lassen Mädchen deutlich bessere Ergebnisse als Jungen erzielen. Diese Befunde korrespondieren mit den Ergebnissen zu den Treatmentunterschieden bei *KogO*. Die einzigen statistisch bedeutsamen treatmentbedingten Differenzen lassen sich auf die sehr gut abschneidenden S-Mädchen zurückführen. In gegenläufiger Weise unterscheiden sich die Geschlechter, Hypothese H_5 wird bezüglich der kognitiven Ebene klar bestätigt.

Tendenziell sind geschlechtsspezifische Treatmentunterschiede auch auf affektiver Ebene zu finden. Hier sollte man die Möglichkeit eines statistischen Irrtums nicht ganz ausschließen. Viele Fälle wurden getestet (Gesamtgruppe, sechs Stationen und diese jeweils in beiden Nachtests) und nur relativ wenige signifikante Unterschiede

gefunden. Eine Bonferroni-Korrektur (vgl. Ergebnisse Kapitel 2.2.2.1.) würde die Signifikanzgrenzen so erheblich erhöhen, dass alle Signifikanzen „geopfert“ würden. Darauf wird verzichtet und eine Interpretation der Ergebnisse trotz der Gefahr eines α -Fehlers versucht (vgl. oben): Wieder spielen die Bekanntheit des Lindenhofs und der Grad an Selbstbestimmung eine entscheidende Rolle. In der Gesamtgruppe bewerten F-Jungen mehrfach Stationen signifikant positiver als die übrigen Treatmentgruppen. Bei Mädchen finden sich keine treatmentbedingten Unterschiede. Zieht man die Bekanntheit mit in Betracht, so entsteht aus den statistisch nachweisbaren affektiven Treatmentunterschieden folgendes Bild:

Tabelle 31: Dargestellt sind im Vergleich zu anderen Treatmentgruppen signifikant positive Einschätzungen von Stationen des Museums durch die Treatmentgruppen von Lindenhof-Nicht-Kennern und Lindenhof-Kennern.

	Nicht-Kenner	Kenner
Jungen	F-Schüler	Keine Unterschiede
Mädchen	F-, öfter SF-Schüler	S-Schüler

Nicht-Kenner fühlen sich anscheinend mit einem relativ fremdbestimmtem Treatment wohl: Das ist für Jungen das F-Treatment, für Mädchen eher das SF-Treatment. S-Mädchen, die den Lindenhof vorher kannten, zeigen verglichen mit den übrigen Treatmentgruppen mehrfach die positivsten affektiven Bewertungen. Damit wird Hypothese H_5 zur Existenz von Geschlechtsunterschiede bei der affektiven Wirkung des Museumsbesuchs gestützt.

Folgende Interpretation bietet sich an: Mädchen bringen als S-Schüler gute kognitive Leistungen (KogO, NT I) und bewerten in relativer Selbstbestimmung (als Nicht-Kenner Treatment SF, als Kenner Treatment S) das Museum besonders positiv. „[Positive] emotionale Beteiligung kann das Lernen erheblich verbessern“ (Spitzer 2002, S. 159 f.). Den Jungen in der Untersuchung liegen dagegen eher geschlossene Antwortformate und ein höheres Maß an Fremdbestimmung.

Man kann vermuten, dass Mädchen eine höhere Ambiguitätstoleranz als Jungen besitzen. Sie bewähren sich in Situationen der Unsicherheit, denn sie schneiden in offenen Tests besser ab, und sie fühlen sich in einem Treatment größerer Selbstbestimmung eher wohl als Jungen.

5. Bedeutung der affektiven Dimension

Nach Cooper (1993) ist an konstruktivistischen Leitlinien orientiertes Lernen intrinsisch motiviert. Die Ergebnisse der vorliegenden Untersuchung zeigen keine höheren affektiven Bewertungen der konstruktivistisch orientierten S-Lerner als der übrigen Lerner. Hypothese H_6 ist damit zu verwerfen.

Die Ursachen dafür könnten in einer Überforderung der S-Schüler zu sehen sein. „lower-ability learners perform better in well-structured (...) instructional environments“ (Cooper 1993, S. 13). Neben der messbaren schlechteren Leistung (vgl. Mandl et al. 1992) von „lower-ability learners“ in Lernsituationen, die ihnen komplexe Lernumgebungen bieten, ist gut vorstellbar, dass schwache Schüler im S-Treatment eine subjektive Überforderung wahrnehmen. Ist dies gepaart mit einem Bewusstsein um den instruktionalen schulischen Rahmen und der damit einhergehenden Erwartung kontrollierender Maßnahmen (vgl. Deci & Ryan 1993), wird eine weniger positive affektive Einschätzung durch diese Probanden offensichtlich. Der Effekt könnte durch eine erhöhte Wahrnehmung von Ambiguität (vgl. Reis 1997), verursacht durch die wahrscheinlich ungewohnteren Aufgaben im S-Treatment, verstärkt werden.

Die implizierte Kausalität der Eingangshypothese lässt sich umkehren: Schüler, die positive affektive Bezüge zu Thema und / oder Lernort zeigen und somit als intrinsisch motiviert gelten können, würden sich eher für das S-Treatment entscheiden und mit diesem höchsten Maß an Selbstbestimmung am besten lernen. Weiter gibt es empirische Befunde, die andeuten, dass die Qualität der Stimmung Einfluss auf die Entscheidung, sich interessegeleiteten Aufgaben zu widmen, haben kann (Krapp 1992, S. 40). Das könnte bedeuten, man sollte versuchen, Schüler bezüglich ihrer Motivationslage (intrinsisch motiviert ?) und ihrer Stimmung einzuschätzen und ihnen je nach dieser Einstufung ein unterschiedliches Maß an Selbstbestimmung zu ermöglichen. Dies sollte in idealer Weise unter der Einbeziehung von Metaebenen erfolgen, denn ein Tun, das eigenen Zielvorstellung entspricht, insbesondere wenn diese Zielvorstellungen bewusst sind, kann zu echter Lernfreude führen (vgl. Krapp 1992, S. 39). Nach der Theorie von Csikszentmihalyi (2000) ist bei intrinsisch motivierten Handlungen besonders häufig das Flow-Erlebnis festzustellen, ein besonderer Zustand, in dem eine Person „voll auf eine Aufgabe konzentriert [ist]“, „(...) ganz in der

Handlung [aufgeht] und sich von äußeren und inneren Störfaktoren kaum beeinflussen lässt. Die Zeit ‚vergeht wie im Flug‘ und die Person hat das sichere Gefühl, den Anforderungen der Situation voll gewachsen zu sein und das Geschehen jederzeit im Blick auf die Handlungsziele optimal kontrollieren zu können“ (Krapp 1992, S. 38). In positiven emotionalen Zuständen kann man Lernaufgaben besser lösen. Man tendiert dazu, ungewöhnliche Aspekte in Überlegungen einzubeziehen und die Grenzen dessen, was man bei der Problemstellung für relevant hält, auszuweiten (Krapp 1992, S. 39). Zum Lernen braucht man „innere Beteiligung“, „eine positive Grundstimmung ist (...) gut für das Lernen (Spitzer 2002, S. 157 ff. und S. 164).

Im instruktionalen Rahmen von Schule herrschen demnach keine besonders günstigen Bedingungen für gutes Lernen. „Eine Lernmotivation, die nicht den Prinzipien des individuellen Selbst entspricht, z.B. weil sie von außen aufoktroziert wird, beeinträchtigt die Effektivität des Lernens (...)“ (Deci & Ryan 1993, S. 235). Deci & Ryan (1993, S. 230) berichten weiter, dass vorhandene intrinsische Motivation durch kontrollierend erlebte Bedingungen zerstört werden kann. Dies können offensichtlich negative Einflüsse wie z.B. Strafandrohung und aufgezwungene Ziele sein, aber auch Termindruck, Bewertungen, Auszeichnungen und materielle Belohnung. Das kann beispielsweise bedeuten: Gibt man einer Person eine Belohnung (z.B. Geld) für etwas, das sie ohnehin tun würde, weil sie es gerne tut, wird sie es danach nicht mehr gerne sondern für die Belohnung (z.B. Geld) tun. Das Tun war vorher autotelischer Natur, intrinsisch motiviert und hat sich durch die kontrollierend erlebte Bedingung zu einem Tun aus extrinsischer Motiviertheit gewandelt. Für die Schule bedeutet dies, dass eine ohne diesen instruktionalen Rahmen intrinsisch motivierte Beschäftigung zu einem Tun als Mittel zum Zweck, z.B. um gute Noten zu bekommen, werden kann. Deci & Ryan (1993, S. 235) folgern, „dass benotete Leistungsprüfungen in der Schule, als die am weitesten verbreiteten Mittel zur Kontrolle der Lernmotivation, ‚Schüsse in den Ofen‘ sind. Sie rufen nicht nur negative affektive Reaktionen hervor, sondern bewirken darüber hinaus auch ein qualitativ schlechteres Lernverhalten.“

Ungelöst bleibt das Problem, von Schülern als uninteressant, langweilig oder mühevoll erlebte Inhalte - die aber als intersubjektive Übereinkunft der Gesellschaft, als zu tradierendes Kulturgut festgesetzt werden - „in die Köpfe der Schüler“ zu bringen. Die meisten Schulsysteme begegnen diesem Problem durch instruktionale Rahmen, die kontrollierende Bedingungen wie Leistungstests mit Bewertungen einschließen.

Emotional positive und gleichzeitig kognitiv sehr produktive Zustände wie ein Flow-Erleben ist in der vorliegenden Untersuchung sehr unwahrscheinlich, aufgrund der zeitlich-organisatorischen Vorgaben, durch die z.B. der instruktionale Rahmen deutlich wird, fast auszuschließen. Ziel sollte es dennoch sein, intrinsische Motivation und entsprechende positive emotionale Zustände zuzulassen. Unterricht, der Selbstbestimmung zulässt, könnte Schülern diese Möglichkeit geben.

6. Unterrichtliche Konsequenzen

Die zentrale Frage nach positiver Attribuierung des Museums und möglichst hohem Lernerfolg durch einen Museumsbesuch aufgreifend, kann über unterrichtliche Konsequenzen spekuliert werden. Die Studie fand im Kontext schulischer Bedingungen statt, so dass Vorbehalte aufgrund eines artifiziellen untersuchungsbedingten Rahmens weitgehend auszuschließen sind. Eine Umsetzung dieser auf quantitativen statistischen Methoden beruhenden Aussagen auf das Unterrichten von Individuen sollte behutsam und mit Augenmaß erfolgen.

Der Museumsbesuch war erfolgreich, affektive Wirkungen sehr positiv, (wenngleich vom Treatment fast unabhängig), kognitive Lernfortschritte persistent. Man kann daher den „Museumsunterricht“ selbst als einmalige halbtägige Intervention in einem wenig „spektakulären“ außerschulischen Lernort wie dem Naturkundemuseum des Umweltschutz-Informationszentrums Lindenhof in Bayreuth empfehlen. Die Tatsache, ob man diesen Unterricht durchführt, spielt eine entscheidendere Rolle, als die, wie man ihn gestaltet. Bezogen auf den Faktor der Selbstbestimmung der Schüler, sollte man keine zu deutlichen instruktionalen Bedingungen schaffen. Die Behaltensleistung wird dadurch negativ beeinflusst. Reale Entscheidungsmöglichkeiten, aber auch subjektiv wahrgenommene Autonomie beeinflussen dagegen Lernprozesse positiv. Überraschenderweise lernen Schüler, die den außerschulischen Lernort

schon kannten, nicht schlechter als die, denen der Lindenhof unbekannt war. Die vielfach zu beobachtende Entscheidung von Lehrern, eine durch einen einmaligen Besuch bekannte Einrichtung nicht wiederzubesuchen, könnte demnach grundlegend falsch sein.

Bezüglich der Lernbedingung Selbstbestimmung ist das Ziel von Unterricht entscheidend. Dies gilt sicher nicht nur für Naturkundemuseen sondern auch für den Lernort Schule. Sollen planbar viele Inhalte vermittelt werden, die nach kurzer Zeit im zur Lernsituation ähnlichen Kontext abzuprüfen sind, führen kleinschrittige instruktionale Vorgaben zu recht guten kognitiven Erfolgen. Die schwachen Behaltensleistungen dieses fremdbestimmten Unterrichts bleiben unbemerkt. Diese Situation ist dem vielfach stattfindenden Biologieunterricht mit seiner Stofffülle und den häufigen Prüfungen über Inhalte der letzten Unterrichtsstunde nicht unähnlich. Unterricht, der Lernen höherer kognitiver Anforderungsstufen entspricht, der „(...) weniger Reproduktion (...) dafür aber verstärkt selbst organisiertes Lernen sowie (...) Anwendung von Wissen in den Vordergrund (...)“ stellt (Klautke 2003, S. 33 ff.), sollte sich eher an Prinzipien einer optimalen Balance zwischen Selbst- und Fremdbestimmung bzw. Konstruktion und Instruktion im Sinne des gemäßigten Konstruktivismus nach Reinmann-Rothmeier & Mandl (2001) halten. Ganz entscheidend ist es, interindividuelle Unterschiede der Schüler zu berücksichtigen. Darum ist eine präzise Positionierung auf diesem Kontinuum zwischen Selbst- und Fremdbestimmung bzw. Konstruktion und Instruktion sicher kaum möglich. Das gilt auch, weil abwechselnde Inhalte und ungleiche Unterrichtssituationen in diesem Punkt wahrscheinlich zu unterschiedlichen Resultaten führen.

Die Entscheidung für diese Positionierung liegt nicht nur bei den Lehrpersonen, sondern auch bei den Intentionen des Lehrplans. Neuere Lehrpläne orientieren sich zwar verstärkt daran, didaktische Absichten zu formulieren. Es bleibt jedoch die Tendenz, möglichst viel Wissen zu vermitteln. Eine Alternative könnte sein, sich mehr auf Grundprinzipien des Fachs zu beschränken und größeres Gewicht auf die jeweiligen fachgemäßen Arbeitsweisen und Methoden des Erkenntniserwerbs zu legen. Hierdurch ließen sich vermehrt handlungsorientierte und selbstbestimmte Lernprozesse initiieren und vielleicht effektiver als bisher neues Wissen und Erkennen in vorhandene subjektive Wissens- und Denkstrukturen des Lernenden einbetten. Nachhaltiges Lernen könnte die Folge sein.

VII. Zusammenfassung

1. Auf der Erkenntnistheorie des Konstruktivismus basierend, existieren unterschiedliche Lerntheorien des Konstruktivismus. Nach der Position des gemäßigten Konstruktivismus von Reinmann-Rothmeier & Mandl (2001) wird das „Prozessmerkmal“ Selbstbestimmung als besonders wesentlich identifiziert und in einer empirischen Studie kontrolliert variiert. Die Untersuchung findet in einem außerschulischen Lernort, einem Naturkundemuseum, das beschrieben und in seiner didaktischen Bedeutung (einschließlich Lernziele) charakterisiert wird, statt. An 366 Gymnasiasten der fünften Jahrgangsstufe wird auf kognitiver und affektiver Ebene die Lernwirkung dieses Faktors in einer eintägigen Intervention evaluiert. Die Schüler durchlaufen im „Museumsunterricht“ entweder ein weitgehend selbstbestimmtes (S-Gruppe), ein fremdbestimmtes (F-Gruppe) Treatment oder eine Mischform aus Selbst- und Fremdbestimmung (SF-Gruppe).

2. In einem Vortest-Nachtest-Design mit Follow-up-Test nach 40 Tagen (Vortest, Nachtest I und II) werden die kognitiven Leistungen der Probanden mit zwei Messinstrumenten, einem Haupttest aus Items mit Antwortvorgaben (Multiple-Choice Test) und einem Nebentest mit Items ohne Antwortvorgaben (offener Test) bestimmt. Die affektiven Bewertungen der Schüler werden in den Nachtests ebenfalls mit geschlossenen und offenen Formaten erfasst.

3. Die zentralen Fragen nach der affektiven Wirkung und dem kognitiven Lernerfolg sind differenziert zu beantworten:

a) Auf affektiver Ebene unterscheiden sich die Schülereinschätzungen der drei Treatments unabhängig vom Messinstrument kaum. In allen Gruppen (Gesamt- und Treatmentgruppen) sind die Bewertungen zu Museumsbesuch insgesamt und Elementen dieses Unterrichts im Nachtest I sehr positiv, im Nachtest II kaum geringer.

b) Auf kognitiver Ebene zeigen beide Messinstrumente für die Gesamtgruppe aller Schüler Erwerb, Persistenz aber auch Vergessen von Wissen an. Der Museumsbesuch führt somit zu bleibendem Lernerfolg.

Bei der Betrachtung der Treatmentgruppen unterscheiden sich die Ergebnisse der beiden Messinstrumente grundlegend:

- Im offenen kognitiven Test sind die Schüler des selbstgesteuerten Treatments am besten. Die beiden anderen Gruppen fallen dagegen deutlich ab, unterscheiden sich untereinander jedoch nicht. Im Nachtest II verschwinden alle Unterschiede zwischen den Treatmentgruppen.
- Im Multiple-Choice Test führt das mittlere Treatment in beiden Nachtests zu den besten Resultaten. Die fremdbestimmten Probanden zeigen nur im Nachtest I akzeptablen Lernerfolg, im Nachtest II dagegen schlechtere Ergebnisse. Auch das selbstgesteuerte Treatment schneidet in beiden Nachtests relativ schwach ab. Eine entsprechende Analyse deutet darauf hin, dass hier eine Beeinträchtigung dieser Treatmentergebnisse resultierend aus Itemkonstruktion und einem der Grundfrage der Untersuchung immanenten Dilemma zurückzuführen sein könnte.

c) Entscheidende Auswirkung hat der Faktor Bekanntheit des Lindenhofs auf die Lernergebnisse des Multiple-Choice Tests. Obwohl die Schüler, die den Lindenhof schon vor diesem Museumsbesuch kannten, von einem höheren Kenntnisstand starten, ist das Niveau des Wissenszuwachses von Kennern und Nicht-Kennern in beiden Nachtests gleich hoch. Die Lindenhof-Nicht-Kenner reagieren in allen entscheidenden Punkten wie die Gesamtgruppe, die Kenner dagegen lassen kein Vergessen nachweisen, nicht in der Gruppe aller Kenner und nicht in den entsprechenden Gruppen der S-, SF- und F- Schüler. Besonders verwunderlich ist ein Fehlen von Treatmentunterschieden. Möglicherweise verfügen Lindenhof-Kenner über andere Strategien, sich das Museum zu erschließen, und werden darum von den Treatmentaufgaben weniger beeinflusst.

4. Unterrichtliche Konsequenzen der Untersuchung:

- a) Exkursionen zu einem Naturkundemuseum lohnen sich. Unabhängig von der Treatmentaufgabe führt der Unterrichtsbesuch zu dauerhaftem Lernerfolg und zu anhaltender positiver affektiver Gestimmtheit bezüglich dieses Besuchs.
- b) Je nach Absicht, sollte man für den Museumsbesuch ein passendes Treatment wählen: Soll es um planbaren Wissenserwerb gehen, einem Unterricht nach operationalisierbaren vorher festgelegten Anforderungen, so ist eine Mischform zwischen Selbstbestimmung und Fremdbestimmung zu empfehlen. Ein sehr deutlich gelenktes Vorgehen kann zu schlechten Behaltensleistungen führen. Ist die Vorbedingung der Planbarkeit verzichtbar, so kann selbstgesteuertes Lernen zu guten Ergebnissen führen und ist bedingt zu empfehlen.
- c) Die vorherige Bekanntheit des Museums bei Schülern sollte nicht von einem weiteren Besuch abhalten. Gerade in dieser Situation lernen Schüler besonders gut. Das Treatment spielt für diese Situation eine untergeordnete Rolle.



Abbildung 62: Der Museumsbesuch endet für die meisten Schüler im Kindermuseum.

VIII. Literatur

Allerbeck, K. R. (1978). Messniveau und Analyseverfahren - Das Problem "strittiger Intervallskalen". *Zeitschrift für Soziologie*, 7(3), 199-214.

Anderson, L. W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Anderson, L. W. & Sosniak, L. A. (1994). *Bloom's Taxonomy. A Forty-year Retrospective* (Vol. 2). Chicago, Illinois: The National Society for the Study of Education.

Apel, H. J. & Knoll, M. (2001). *Aus Projekten lernen. Grundlegung und Anregungen*. München: Oldenbourg Schulbuchverlag.

Ausubel, D. P. (1974). *Psychologie des Unterrichts* (Vol. 1). Weinheim: Beltz.

Baptist, P. (2003). Eigene Lernwege gehen. *Lernchancen*, 31, 42-45.

Bauer, H.-G. & Berthold, P. (1997). *Die Brutvögel Mitteleuropas* (2. ed.). Wiesbaden: Aula-Verlag.

Bauer, E. W., Herzinger, H., Kleesattel, W., Reichelt, G., & Schwoerbel, W. (1988). *Biologie für bayerische Gymnasien Jahrgangsstufe 9*. Berlin: Cornelsen Verlag.

Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Eds.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (2002). *PISA 2000 - Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.

Bay, F. & Rodi, D. (1991). Beziehungen zwischen Beutegreifern und Beutetieren. In D. Eschenhagen & U. Kattmann & D. Rodi (Eds.), *Handbuch des Biologieunterrichts. Sekundarbereich I* (Vol. 8. Umwelt, pp. 32-48). Köln: Aulis Verlag.

Bay, F., Schneider, H. & Strecker, H. (1993). *Natura Lehrerband Teil A* (Vol. 2). Stuttgart: Ernst Klett Schulbuchverlag.

Becker, G. (2000). *Vom ökologischen Lernen zur Bildung für eine nachhaltige Entwicklung*. Osnabrück: Universitätsverlag Rasch.

Benninghaus, H. (1991). *Einführung in die sozialwissenschaftliche Datenanalyse* (2. ed.). München: Oldenbourg.

Berck, K.-H. (1999). *Biologiedidaktik. Grundlagen und Methoden*. Wiebelsheim: Quelle & Meyer Verlag.

Berck, K.-H. & Graf, D. (Eds.). (1987). *Rahmenplan des Verbandes Deutscher Biologen für das Schulfach Biologie*. Publikation Nr. 7 des Verbandes Deutscher Biologen e.V. Bremen: VDBiol

Berck, K.-H. & Graf, D. (2003). *Biologiedidaktik von A bis Z. Wörterbuch mit 1000 Begriffen*. Wiebelsheim: Quelle & Meyer Verlag.

Berndt, R. (1970). Meisen - Kleiber und Baumläufer - Familie Eigentliche Meisen. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 9. Vögel 3, pp. 298-304). Zürich: Kindler Verlag.

Beyer, H. & Walter, W. (1988). *Lehrbuch der Organischen Chemie* (21. ed.). Stuttgart: Hirzel Verlag.

Bezzel, E. (1982). *Vögel in der Kulturlandschaft*. Stuttgart: Verlag Eugen Ulmer.

Bezzel, E. (1985). *Kompendium der Vögel Mitteleuropas. Nonpasseriformes - Nichtsingvögel*. Wiesbaden: Aula-Verlag.

Bezzel, E. & Prinzinger, R. (1990). *Ornithologie* (2. ed.). Stuttgart: Verlag Eugen Ulmer.

Bickel, H., Claus, R., Frank, R., Gropengießer, H., Haala, G., Knauer, B., Kronberg, I., Lichtner, H.-D., Loth, U., Schweizer, J., Sommermann, U., Ströhla, G., Tischler, W. & Wichert, G. (1998). *Natura Oberstufe* (Vol. 3). Stuttgart: Ernst Klett Schulbuchverlag.

Bleymüller, J., Gehlert, G. & Gülicher, H. (2000). *Statistik für Wirtschaftswissenschaftler*. München: Verlag Franz Vahlen.

Bloom, B. S., Engelhart, M. D., Fürst, E. J., Hill, W. H. & Krathwohl, D. R. (1976). *Taxonomie von Lernzielen im kognitiven Bereich* (5. ed.). Weinheim und Basel: Beltz Verlag.

Bogner, F. X. (1998). The Influence of Short-Term Outdoor Ecology Education on Long-Term Variables of Environmental Perspective. *The Journal of Environmental Education*, 29(4), 17-29.

Bogner, F. X. & Wiseman, M. (2004). Outdoor Ecology Education and Pupils' Environmental Perception in Preservation and Utilization. *Science Education International*, 15(1), 27-48.

Bonora, V., Dieterle, A., Görz, G. & Gotzler, H. (1992 a). *Natura 5. Jahrgangsstufe*. Stuttgart: Ernst Klett Schulbuchverlag.

Bonora, V., Dieterle, A., Görz, G. & Gotzler, H. (1992 b). *Natura 6. Jahrgangsstufe*. Stuttgart: Ernst Klett Schulbuchverlag.

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. ed.). Heidelberg, Berlin, New York: Springer-Verlag.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Heidelberg, Berlin, New York: Springer Verlag.

Brohmer, P., Tischler, W. & Schaefer, M. (1984). *Fauna von Deutschland. Ein Bestimmungsbuch unserer heimischen Tierwelt* (16. ed.). Heidelberg: Quelle & Meyer.

Brown, J. H. (1995). *Macroecology*. Chicago: University of Chicago Press.

Bruner, J. S. (1967). *Towards a Theory of Instruction*. Cambridge Massachusetts: The Belknap Press of Harvard University Press.

Bruner, J. S. (1973). Der Akt der Entdeckung. In H. Neber (Ed.), *Entdeckendes Lernen*. Weinheim und Basel: Beltz Verlag.

Busch, W. (1959). *Und die Moral von der Geschicht* (Vol. 1). München: Bertelsmann.

Büschges-Abel, W. (1998). *Lernen wird zum Erlebnis. Neurolinguistisches Programmieren in Schule und Sozialpädagogik*. Neuwied, Kriftel, Berlin: Hermann Luchterhand Verlag.

Carlson, R., Chandler, P. & Sweller, J. (2003). Learning and Understanding Science. Instructional Material. *Journal of Educational Psychology*, 95(3), 629-640.

Claus, R. & Haala, G. (Eds.). (1987). *Arbeitsblätter Vögel*. Stuttgart: Ernst Klett Verlag.

Claus, R., Dobler, H.-J., Frank, R., Haala, G., Lauer, V., Mura, B., Stephan, F., Strecker, H. & Wichert, G. (1989). *Natura 5. und 6. Schuljahr* (Vol. 1). Stuttgart: Ernst Klett Schulbuchverlag.

Cohen, J. (1968). Weighted Kappa: Nominale Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, 70(4), 213-220.

Cooper, P. A. (1993). Paradigm Shifts in Designed Instruction: From Behaviorism to Cognitivism to Constructivism. *Educational Technology*, 33(5), 12-19.

Corbet, G. & Ovenden, D. (1982). *Pareys Buch der Säugetiere. Alle wildlebenden Säugetiere Europas*. Hamburg, Berlin: Verlag Paul Parey.

Csikszentmihalyi, M. (2000). *Das flow-Erlebnis. Jenseits von Angst und Langeweile: im Tun aufgehen* (8. ed.). Stuttgart: Klett-Cotta.

Daumer, K., Eickenbusch, J. & Libera, W. (1981). *Biologie für den Sekundarbereich I* (Vol. 7 G). München: Bayerischer Schulbuch-Verlag.

Daumer, K. & Schuster, M. (Eds.). (1983). *Stoffwechsel, Ökologie und Umweltschutz*. München: Bayerischer Schulbuch-Verlag.

Daumer, K. (1985). *Biologie für den Sekundarbereich I* (Vol. 9 G). München: Bayerischer Schulbuch-Verlag.

Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York and London: Plenum Press.

Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223-238.

Deci, E. L. & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227-268.

Devitt, M. (1991). Realism & Truth. Zitiert nach M. R. Matthews (Ed.), *Constructivism in Science Education. A Philosophical Examination* (2. ed.). Dordrecht, Boston, London: Kluwer Academic Publishers.

Dewey, J. (1916). *Democracy and Education*. In K. Frey (Ed.), *Die Projektmethode*. Weinheim und Basel: Beltz Verlag.

Diehl, J. M. & Staufenbiel, T. (2002). *Statistik mit SPSS - Version 10+11*. Frankfurt am Main: Klotz Verlag.

Dierl, W. (1969). Die Schmetterlinge. In B. Grzimek (Ed.), *Grzimeks Tierleben Enzyklopädie des Tierreiches* (Vol. 2 Insekten, pp. 306-370). Zürich: Kindler Verlag.

Diesener, G. & Reichholf, J. (1986). *Lurche und Kriechtiere*. München: Mosaik Verlag.

Dorst, J. (1970). Fliegenschnäpper und Drosseln - Unterfamilie Drosseln. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 9. Vögel 3, pp. 271-274). Zürich: Kindler Verlag.

Driver, R. (1995). Constructivist Approaches to Science Teaching. In L. P. Steffe & J. Gale (Eds.), *Constructivism in Education* (pp. 385-400). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Duit, R. *Bibliography - STCSE. Students' and Teachers' Conceptions and Science Education*. Zugriff am 28.05.2004 unter <http://www.ipn.uni-kiel.de/aktuell/stcse/>

Ellenberg, H. (1996). *Vegetation Mitteleuropas mit den Alpen in ökologischer, dynamischer und historischer Sicht* (5. ed.). Stuttgart: Eugen Ulmer.

Eschenhagen, D., Kattmann, U. & Rodi, D. (2001). *Fachdidaktik Biologie* (5. ed.). Köln: Aulis Verlag.

Fechter, R. & Falkner, G. (1989). *Weichtiere*. München: Mosaik Verlag.

Felser, G. (2001). *Werbe- und Konsumentenpsychologie* (2. ed.). Heidelberg, Berlin; Stuttgart: Schäffer-Pöschel.

Fischer, F., Gräsel, C., Kittel, A. & Mandl, H. (1997). Strategien zur Bearbeitung von Diagnoseproblemen in komplexen Lernumgebungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, XXIX(1), 62-82.

Fischer, H., Klemm, K., Leutner, D., Sumfleth, E., Tiemann, R. & Wirth, J. (2003). Naturwissenschaftsdidaktische Lehr- Lernforschung: Defizite und Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 9. Jg., 179-209.

Fischler, H. (1999). Sammelrezension: Konstruktivismus in den Didaktiken der Mathematik und der Naturwissenschaften. *Zeitschrift für Erziehungswissenschaft*, 2(4), 569-578.

Foerster, H. v. (2001). Das Konstruieren einer Wirklichkeit. In P. Watzlawick (Ed.), *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus* (13. ed., pp. 326). München: Piper Verlag.

Frank, F. (2001). Das Geowissenschaftliche Museum als außerschulischer Lernort - Überlegungen und Untersuchungen am Beispiel des Riesenkrater-Museums Nördlingen. *Münchner Studien zur Didaktik der Geographie*, 11.

Friedrich, G. & Preiss, G. (2002). Lehren mit Köpfchen. *Gehirn & Geist*, Jg. 2002(4), 64-70.

Frey, K. (1982). *Die Projekt-Methode*. Weinheim und Basel: Beltz Verlag.

Freytag, G. E. (1970). Schwanzlurche und Blindwühlen. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 5. Fische 2. Lurche, pp. 313-358). Zürich: Kindler Verlag.

Gebhard, U. (1990). Dürfen Kinder Naturphänomene beseelen? *Unterricht Biologie*, 153, 38-42.

Gerstenmaier, J. & Mandl, H. (1995). Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik*, 41(6), 867-888.

Glaserfeld, E. v. (1987). *Wissen, Sprache und Wirklichkeit. Arbeiten zum radikalen Konstruktivismus*. Braunschweig: Friedrich Vieweg & Sohn.

Glaserfeld, E. v. (1993). Questions and Answers about Radical Constructivism. In K. Tobin (Ed.), *The Practice of Constructivism in Science Education* (pp. 23-38). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Glaserfeld, E. v. (1996). *Radikaler Konstruktivismus. Ideen, Ergebnisse, Probleme*. Frankfurt am Main: Suhrkamp.

Glaserfeld, E. v. (1998). *Konstruktivismus statt Erkenntnistheorie*. Klagensfurt: Drava Verlag.

Glaserfeld, E. v. (2001). Einführung in den radikalen Konstruktivismus. In P. Watzlawick (Ed.), *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus*. (13. ed., pp. 16-38). München: Piper.

Grant, S. & Zetterström, M. (2000). *Vögel Europas, Nordafrikas und Vorderasiens*. Stuttgart: Kosmos.

Gräsel, C. (1997). *Problemorientiertes Lernen*. Göttingen: Hogrefe Verlag für Psychologie.

Gries, B. (1996). Das Naturkundemuseum als außerschulischer Lernort. In H. Vogt & M. Sieger (Eds.), *Berichte des Instituts für Didaktik der Biologie der Westfälischen Wilhelms-Universität Münster* 5 (1996), 1-18.

Grolnick, W. S. & Ryan, R. M. (1987). Autonomy in Children's Learning: An Experimental and Individual Difference Investigation. *Journal of Personality and Social Psychology*, 52(5), 890-898.

Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Eds.), *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Handlungsansätze*. Göttingen: Hogrefe Verlag für Psychologie.

Gubitz, C. & Pfeifer, R. (1993). Die Vogelwelt Ostoberfrankens. *Beihefte zu Berichten der Naturwissenschaftlichen Gesellschaft Bayreuth*, 3.

Gunstone, R. F. (2000). Constructivism and Learning Research in Science Education, *Constructivism in Education* (pp. 254-280). Chicago, Illinois.

Haase, H.-M. & Bogner, F. X. (2002). Nachhaltigkeit und Umweltbildung: Möglichkeiten Pädagogischen Wirkens. *Natur und Kultur*, 3(2), 75-94.

Hacker, H. (1979). *Elemente des Curriculums*. Donauwörth: Verlag Ludwig Auer.

Hansmann, O. (1998). *Operative Pädagogik. Anlässe zur Reflexion für die Lehrberufe*. Weinheim: Deutscher Studien Verlag.

Hänze, M. (1998). *Denken und Gefühl. Wechselwirkung von Emotion und Kognition im Unterricht*. Neuwied, Kriftel, Berlin: Hermann Luchterhand Verlag.

Hascher, T. (1994). *Emotionsbeschreibung und Emotionsverstehen. Zur Entwicklung des Emotionsvokabulars und des Ambivalenzverstehens im Kindesalter*. Münster, New York: Waxmann Verlag.

Häußler, P., Bünder, W., Duit, R., Gräber, W. & Mayer, J. (1998). *Naturwissenschaftsdidaktische Forschung. Perspektiven für die Unterrichtspraxis*. Kiel: Institut für die Pädagogik der Naturwissenschaften.

Hay, K. E. & Barab, S. A. (2001). Constructivism in Practice: A Comparison and Contrast of Apprenticeship and Constructionist Learning Environments. *The Journal of the Learning Sciences*, 10(3), 281-322.

Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P. (Eds.). (1985). *The Beginnings of birds*. Willibaldsburg: Freunde des Jura-Museums Eichstätt.

Hoops, W. (1998). Konstruktivismus. Ein neues Paradigma für Didaktisches Design? *Unterrichtswissenschaft*, 26(3), 229-253.

Horster, L. (2002). Die Entwicklung eines neuen Verständnisses von Lernen nach PISA. Eine gemeinsame Aufgabe von Schulleitung und Kollegium. In H. Buchen & L. Horster & G. Pantel & H.-G. Rolff (Eds.), *Unterrichtsentwicklung nach PISA*. Berlin: RAABE-Fachverlag für Bildungsmanagement.

Immelmann, K. (1982). *Wörterbuch der Verhaltensforschung*. Berlin, Hamburg: Verlag Paul Parey.

Issing, L. J. (1997). Instruktionsdesign für Multimedia. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (2. ed.). Weinheim: Beltz PVU.

Issing, L. J. & Klimsa, P. (1997). *Information und Lernen mit Multimedia* (2. ed.). Weinheim: Beltz PVU.

Jacobs, W. & Renner, M. (1988). *Biologie und Ökologie der Insekten. Ein Taschenlexikon* (2. ed.). Stuttgart: Gustav Fischer Verlag.

Jungbauer, W. & Weber, W. (Eds.). (2001). *Netzwerk Biologie 2. Baden-Württemberg*. Hannover: Schroedel Verlag.

Kaiser, F. G. & Fuhrer, U. (2000). Wissen für ökologisches Handeln. In H. Mandl & J. Gerstenmaier (Eds.), *Die Kluft zwischen Wissen und Handeln*. Göttingen: Hogrefe Verlag für Psychologie.

Kalusche, D. (1999). *Ökologie - ein Lehrbuch* (3. ed. Vol. 25). Wiesbaden: Quelle & Meyer.

Kerschensteiner, G. (1921). *Grundfragen der Schulorganisation* (4. ed.). Leipzig und Berlin: B. G. Teubner.

Kerstan, T. & Spiewak, M. (2003). Die Bilanz des Schreckens. *Die Zeit*, 50. Zugriff am 26.05.2004 unter: <http://www.pisaoecd.org>

Killermann, W. (1995). *Biologieunterricht heute* (10. ed.). Donauwörth: Verlag Ludwig Auer.

Killermann, W. (2000). Ganzheitliche Naturschutz- und Umwelterziehung. *Berichte der Bayerischen Akademie für Naturschutz und Landschaftspflege*, 24, 21-26.

Klafki, W. (1964). Didaktische Analyse als Kern der Unterrichtsvorbereitung. In H. Roth & A. Blumenthal (Eds.), *Auswahl 1; Grundlegende Aufsätze aus der Zeitschrift Die Deutsche Schule* (pp. 5-34). Hannover: Schroedel.

Klafki, W. (1980). Zur Unterrichtsplanung im Sinne kritisch-konstruktiver Didaktik. In E. König & N. Schier & U. Vohland (Eds.), *Diskussion Unterrichtsvorbereitung. Verfahren und Modelle*. München: Wilhelm Fink Verlag.

Klautke, S. (2003). Unterrichtsverfahren im Biologieunterricht. Entdeckend- Forschendes Lernen. *Lernchancen*, 31.

König, C. (1969). Die Eulen. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 8. Vögel 2, pp. 377-406). Zürich: Kindler Verlag.

König, E. & Riedel, H. (1971). *Unterrichtsplanung als Konstruktion* (2. ed.). Weinheim: Verlag Julius Beltz.

König, G. (1999). *Kinder- und Jugendmuseen. Genese und Entwicklung einer Museumsgattung. Impulse für ein Besucherorientiertes Museumskonzept*. Dissertation, Universität Tübingen.

Krapp, A. (1992). Konzepte und Forschungsansätze zur Analyse des Zusammenhangs von Interesse, Lernen und Leistung. In A. Krapp & M. Prenzel (Eds.), *Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (pp. 9-52). Münster: Aschendorff.

Krathwohl, D. R., Bloom, B. S. & Masia, B. B. (1975). *Taxonomie von Lernzielen im affektiven Bereich*. Weinheim und Basel: Beltz Verlag.

Kudrna, O. (2002). The Distribution Atlas of European Butterflies. *Oedippus*, 20, 1-342.

Küster, H. (1995). *Geschichte der Landschaft in Mitteleuropa. Von der Eiszeit bis zur Gegenwart*. München: Verlag C. H. Beck.

Ladenthin, V. (2001). Die Frage nach der Selbstdarstellung von Wissen. *Forschung und Lehre*, 6, 306-308.

Lamprecht, J. (1999). *Biologische Forschung. Von der Planung bis zur Publikation*. Fürth: Filander Verlag.

Lantermann, E. D., Döring-Seipel, E. & Schima, P. (1992). Werte, Gefühle und Unbestimmtheit: kognitiv-emotionale Wechselwirkungen im Umgang mit einem ökologischen System. Zitiert nach H.-M. Haase & F. X. Bogner (Ed.), *Nachhaltigkeit und Umweltbildung: Möglichkeiten pädagogischen Wirkens. Natur und Kultur* 3/2, 75-94

Law, L.-C. (2000). Die Überwindung der Kluft zwischen Wissen und Handeln aus situativer Sicht. In H. Mandl & J. Gerstenmaier (Eds.), *Die Kluft zwischen Wissen und Handeln*. Göttingen: Hogrefe Verlag für Psychologie.

Leibhold, K. (1997). *Modelle, Modellbildung und Modelleinsatz - untersucht am Beispiel des Energiestoffwechsels der Pflanzen für den Biologieleistungskurs des Gymnasiums*. Inaugural-Dissertation, Universität Bayreuth.

Leinhardt, G. (1993). On Teaching. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 1-54). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lienert, G. A. (1969). *Testaufbau und Testanalyse* (3. ed.). Weinheim: Beltz Verlag.

Lionni, L. (1970). *Fish is Fish*. New York: Dragonfly books.

Lodder, C. (1985). *Russian Constructivism*. New Haven and London: Yale University Press.

Lord, T. R. (1997). A Comparison Between Traditional and Constructivist Teaching in College Biology. *Innovative Higher Education*, 21(3), 197-216.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & D'Appolonia, S. (1996). Within-class grouping: a meta-analysis. *Review of Educational Research*, 66, 423-458.

Lucke, D. (1995). *Akzeptanz. Legitimität in der "Abstimmungsgesellschaft"*. Opladen: Leske + Budrich.

Lude, A. (2001). *Naturerfahrung und Naturschutzbewusstsein. Eine empirische Studie* Reihe Forschungen zur Fachdidaktik (Vol. 2). Innsbruck: StudienVerlag.

Mager, R. F. (1965). *Lernziele und programmierter Unterricht*. Weinheim: Beltz Verlag.

Madl, W. & Bauer, R. (1982). *Biologie für den Sekundarbereich I* (Vol. 8 G). München: Bayerischer Schulbuch-Verlag.

Mandl, H. (2001). *Die Entwicklung einer neuen Lehr- und Lernkultur*. Paper presented at the 19. Lehrertag Oberfranken in Bayreuth.

Mandl, H., Gruber, H. & Renkl, A. (1997). Situiertes Lernen in Multimedialen Lernumgebungen. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (2. ed., pp. 168-178). Weinheim: Beltz PVU.

Mandl, H., Schnotz, W., Picard, E. & Henninger, M. (1992). Knowledge Acquisition with Texts by Means of Flexible Computer-Assisted Information Access. In A. Oliveira (Ed.), *Hypermedia Courseware* (pp. 70-76). Berlin: Springer Verlag.

Matthews, M. R. (1998). Introductory Comments on Philosophy and Constructivism in Science Education. In M. R. Matthews (Ed.), *Constructivism in Science Education. A Philosophical Examination* (pp. 1-10). Dordrecht, Boston, London: Kluwer Academic Publishers.

Maturana, H. R. & Varela, F. (1987). *Der Baum der Erkenntnis. Die biologischen Wurzeln des menschlichen Erkennens*. Bern, München, Wien: Scherz Verlag.

Mayer, J. (2002). Biologieunterricht nach PISA. Standards, Qualitätsentwicklung und Evaluation des Biologieunterrichts. In H. Buchen & L. Horster & G. Pantel & H.-G. Rolff (Eds.), *Unterrichtsentwicklung nach PISA*. Berlin: RAABE Fachverlag für Bildungsmanagement.

Mitterer, J. (1998). Eine Laudatio von Josef Mitterer. In W. Dörfler & J. Mitterer (Eds.), *Ernst von Glaserfeld - Konstruktivismus statt Erkenntnistheorie* (pp. 109-120). Klagenfurt: Drava Verlag.

Möller, C. (1976). *Technik der Lernplanung* (5. ed.). Weinheim und Basel: Beltz Verlag.

Müller-Using, D. (1968). Rauhfußhühner - Die Haselhühner. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 7. Vögel 1, pp. 459-460). Zürich: Kindler Verlag.

Neber, H. (1973). *Entdeckendes Lernen*. Weinheim und Basel: Beltz Verlag.

Newton, I. (1970). Die Finkenfamilie. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 9. Vögel 3, pp. 387-404). Zürich: Kindler Verlag.

Nitsche, G. & Plachter, H. (1987). *Atlas der Brutvögel Bayerns 1979-1983*. München: Ornithologische Gesellschaft in Bayern.

National Science Teachers Association (NSTA), (1969). *Behavioural Objectives in the Affective Domain*: National Education Association.

Nuissl, E. (1988). Vermittlung im wachsenden Museum. Zitiert nach B. Gries (Ed.), *Das Naturkundemuseum als außerschulischer Lernort. Berichte des Instituts für Didaktik der Biologie der Westfälischen Wilhelms-Universität Münster* 5 (1996), 1-18.

Odum, E. P. & Reichholf, J. (1980). *Ökologie* (4. ed.). München, Wien, Zürich: BLV Verlagsgesellschaft.

Owen, E. & Sweller, J. (1985). What Do Students Learn While Solving Mathematics Problems? *Journal of Educational Psychology*, 77(3), 272-284.

Pekrun, R. (1988). *Emotion, Motivation und Persönlichkeit* (Vol. 1). München und Weinheim: Psychologie Verlags Union.

Peters, R. H. (1983). *The ecological implications of body size*. Cambridge, London, New York: Cambridge University Press.

Piaget, J. (1974). *Der Aufbau der Wirklichkeit beim Kinde*. Stuttgart: Ernst Klett Verlag.

Popper, K. R. (1984). *Logik der Forschung* (8. ed.). Tübingen: Mohr.

Ramachandran, V. S., & Blakeslee, S. (1998). *Die blinde Frau, die sehen kann. Rätselhafte Phänomene unseres Bewußtseins*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.

Randler, C. (2004). *Kognitive und emotionale Faktoren des Lernens am Beispiel einer Biologieunterrichtseinheit "Lebensraum See"*. Dissertation, Pädagogische Hochschule Ludwigsburg: Verlag Dr. Kovac.

Randler, C. & Bogner, F. X. (2002). Comparing Methods of Instruction Using Bird Species Identification Skills as Indicators. *Journal of Biological Education*, 36(4), 181-188.

Reck, M. (2000). Mathematisches Modell zur Symbiose - Modellkonzeption und Evaluation des Modelleinsatzes im Biologieunterricht. In Horst Bayrhuber et al. (Ed.), *Forschungen zur Fachdidaktik* (Vol. 1, pp. 250-262). Innsbruck: Studienverlag.

Reichholf, J. (1980). Die Arten-Areal-Kurve bei Vögeln in Mitteleuropa. *Anzeiger der Ornithologischen Gesellschaft in Bayern*, 19, 13-26.

Reichholf, J. H. (1988). Die Verarmung unserer Umwelt aus der Sicht des Zoologen. *Forstwissenschaftliches Centralblatt*, 107, 263-273.

Reichholf, J. H. (1998). *Der Blaue Planet*. München: Deutscher Taschenbuch Verlag.

Reinmann-Rothmeier, G. & Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie* (4. ed., pp. 601-646). Weinheim und Basel: Beltz PVU.

Reinmann-Rothmeier, G., Mandl, H. & Prenzel, M. (1994). *Computerunterstützte Lernumgebungen*. Erlangen: Publicis MCD Verlag.

Reis, J. (1996). *Inventar zur Messung der Ambiguitätstoleranz (IMA). Manual*. Heidelberg: Roland Asanger Verlag.

Reis, J. (1997). *Ambiguitätstoleranz. Beiträge zur Entwicklung eines Persönlichkeitskonstruktes*. Heidelberg: Roland Asanger Verlag.

Remer, A. (1994). *Organisationslehre* (2. ed.). Hummeltal: R.E.A. Verlag Managementforschung.

Rogoff, B. (1990). *Apprenticeship in Thinking. Cognitive Development in Social Context*. New York, Oxford: Oxford University Press.

Rost, J. (1996). *Lehrbuch Testtheorie und Testkonstruktion*. Bern: Verlag Hans Huber.

Roth, G. (1997). *Das Gehirn und seine Wirklichkeit. Kognitive Wirklichkeit und ihre philosophischen Konsequenzen* (4. ed.). Frankfurt am Main: Suhrkamp Taschenbuch Wissenschaft.

Rowling, J. K. (1999). *Harry Potter and the Prisoner of Azkaban*. London: Bloomsbury Publishing.

Schaefer, M. (1992). *Ökologie* (3. ed.). Jena: Gustav Fischer Verlag.

Scharf, K.-H. & Weiß, K.-D. (1994 a). *Umwelt: Biologie 7. Bayern*. Stuttgart: Ernst Klett Schulbuchverlag.

Scharf, K.-H. & Weiß, K.-D. (1994 b). *Umwelt: Biologie 8. Bayern*. Stuttgart: Ernst Klett Schulbuchverlag.

Scharfenberg, F.-J. & Klautke, S. (2003). Das Demonstrationslabor Bio-/Gentechnik der Universität Bayreuth: Aufbau und Evaluation. In H. Vogt & D. Krüger & U. Unterbrunner (Eds.), *Erkenntnisweg Biologiedidaktik. Beiträge auf der 5. Frühjahrsschule der Sektion Biologiedidaktik im VDBiol in Salzburg 2003* (pp. 25-46).

Schmidt, H. (1979). *Die Wiese als Ökosystem*. Mönchengladbach: Aulis Verlag.

Schmitt-Scheerso, A. & Vogt, H. (2003). Das Naturkundemuseum als Interesse fördernder Lernort - Besucherstudie in einer naturkundlichen Ausstellung. In A. Bauer et al. (Ed.), *Entwicklung von Wissen und Kompetenzen im Biologieunterricht* (pp. 175-182). *Internationale Tagung der Sektion Biologiedidaktik im VDBiol 14.09.-19.09.2003*.

Schorch, G. (1998). *Grundschulpädagogik - eine Einführung*. Bad Heilbrunn: Verlag Julius Klinkhardt.

Schulz-Zander, R. (1997). Kind und Computer: Neue Lernmöglichkeiten in der Schule. In M. Liedke (Ed.), *Kind und Medien. Zur Kulturgeschichtlichen und Ontogenetischen Entwicklung einer Beziehung* (pp. 265-280). Bad Heilbrunn: Verlag Julius Klinkhardt.

Seck-Lanzendorf, S. v. (1997). *Der Einfluss des Ökofaktors Erkrankungen auf die Populationsentwicklung des Feldhasen (Lepus europaeus) im Forschungsrevier Czempin in Polen*. Inaugural-Dissertation, Freie Universität Berlin.

Skinner, B. F. (1971). *Erziehung als Verhaltensformung. Grundlagen einer Technologie des Lehrens*. München-Neubiberg: E. Keimer Verlag.

Sommer, M. (2003). *SOMSO Modelle*. Firmenkatalog. Coburg.

Spitzer, M. (2002). *Lernen. Gehirnforschung und die Schule des Lebens*. Heidelberg: Spektrum Akademischer Verlag.

Spitzer, M. (2004). *Selbstbestimmen. Gehirnforschung und die Frage: Was sollen wir tun?* Heidelberg, Berlin: Spektrum Akademischer Verlag.

Staatsministerium für Unterricht und Kultus (16.06.1983). *Schulordnung für die Gymnasien in Bayern*. Zugriff am 19.04.2004 unter <http://www.km.bayern.de/imperia/md/content/pdf/schulen/gso.pdf>

Staatsministerium für Unterricht und Kultus (1981). *Schulordnung für die Grundschulen in Bayern*. KMBI I So.-Nr. 20/1981.

Staeck, L. (1995). *Zeitgemäßer Biologieunterricht* (5. ed.). Berlin: Cornelsen Verlag.

Suchting, W. (1998). Reflections on Peter Slezak and the "Sociology of Scientific Knowledge". In M. R. Matthews (Ed.), *Constructivism in Science Education. A Philosophical Examination*. Dordrecht, Boston, London: Kluwer Academic Publisher.

Sweller, J., Chandler, P., Tierney, P. & Cooper, M. (1990). Cognitive Load as a Factor in the Structuring of Technical Material. *Journal of Experimental Psychology: General*, 119(2), 176-192.

Terhart, E. (1999). Konstruktivismus und Unterricht. Gibt es einen neuen Ansatz in der Allgemeinen Didaktik? *Zeitschrift für Pädagogik*, 45(5), 629-647.

Tobin, K. (Ed.). (1993). *The Practice of Constructivism in Science Education*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Tompson, T. L. & Mintzes, J. J. (2002). Cognitive Structure and the Affective Domain: on Knowing and Feeling in Biology. *International Journal of Science Education*, 24(6), 645-660.

Tovee, M. J., Rolls, E. T. & Ramachandran, V. S. (1996). Rapid Visual Learning in Neurones of the Primate Temporal Visual Cortex. *NeuroReport*, 7, 2757-2760.

Urhahne, D. & Schanze, S. (2003). Wie lässt sich das Lernen mit Hypertext effektiver gestalten? Empirischer Vergleich einer linearen und einer netzwerkartigen hypermedialen Lernumgebung. *Unterrichtswissenschaft*, 31(4), 359-377.

Uhlig, A. (1962). *Didaktik des Biologieunterrichts*. Berlin: VEB Deutscher Verlag der Wissenschaften.

Watzlawick, P. (2001). *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus* (13. ed.). München: Piper Verlag.

Weid, S. (1995). Wacholderheiden, Schäferei und Landschaftspflege in der Fränkischen Schweiz. *Heimatbeilage zum Amtlichen Schulanzeiger des Regierungsbezirks Oberfranken*, 222.

Weid, S. (1996). Zur Ökologie von Wacholderheiden in der Fränkischen Schweiz. *Heimatbeilage zum Amtlichen Schulanzeiger des Regierungsbezirks Oberfranken*, 233.

Wendt, H. (1970). Die Spechtvögel - Familie Spechte. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 9. Vögel 3, pp. 86). Zürich: Kindler Verlag.

Westphalen, K. (1980). *Praxisnahe Curriculumentwicklung* (8. ed.). Donauwörth: Verlag Ludwig Auer.

White, R. & Gunstone, R. (1992). *Probing Understanding*. London: Falmer Press.

Wilde, M. (2002). Untersuchung zur unterrichtlichen Nutzung eines Naturkundemuseums. In H. Vogt & C. Retzlaff-Fürst (Eds.), *Erkenntnisweg Biologiedidaktik* (pp. 101-111). *Beiträge der 4. Frühjahrsschule der Sektion Biologiedidaktik im VDBiol in Rostock-Warnemünde 2002*.

Wilde, M., Urhahne, D. & Klautke, S. (2003). Unterricht im Naturkundemuseum: Untersuchung über das "richtige" Maß an Instruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 125-134.

Winkler, R. G. (1981). *Werner Gräf und der Konstruktivismus in Deutschland 1918 - 1934*. Dissertation, Aachen

Wolf, B. (2001). Effektstärkenmaße. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (2. ed., pp. 96-102). Weinheim: Beltz PVU.

Wüst, W. (1970). Die Rabenverwandten - Familie Rabenvögel. In B. Grzimek (Ed.), *Grzimeks Tierleben. Enzyklopädie des Tierreiches* (Vol. 9. Vögel 3, pp. 485-506). Zürich: Kindler Verlag.

Zabel, E. (1992). *Pflanzenphysiologie und Ökologie* (Lehrbuch für Klasse 9). Berlin: Volk und Wissen Verlag.

Ziller, T. (1876). Die Theorie der formalen Stufen des Unterrichts. Zitiert nach G. Reinmann-Rothmeier & H. Mandl: Unterrichten und Lernumgebungen gestalten in A. Krapp, B. Weidenmann (Ed.), *Pädagogische Psychologie. Ein Lehrbuch* (4. ed.). 2001: Beltz PVU.

Zöfel, P. (2002). *Statistik verstehen. Ein Begleitbuch zur Computergestützten Anwendung*. München: Addison-Wesley Verlag.



Einleitung:

„Ein Klabautermann versteckt dauernd meine Kappe, Hein Blöd hat das Schiff auf eine Sandbank gesetzt und die drei Bärchen singen seit 17 Stunden.... –Ich brauch’ Urlaub!
„Wacholderheide im Frankenjura“??? – Klingt ja komisch! Was soll das sein?! Wo gib’t’s das denn?“

Wenn Du Gößweinstein, Waischenfeld oder Pottenstein kennst, dann kannst Du Blaubär weiterhelfen. Du warst bestimmt schon einmal da! Das sind alles Orte mitten in der Fränkischen Schweiz! Und genau dort, in der Fränkischen Schweiz nämlich, findet man diese Landschaftsform „Wacholderheide im Frankenjura“ überall.

Aufgabe:

Der Raubwürger ist ein typischer Würger. Finde mit Hilfe der Karte zum Schaufenster den Raubwürger!

Fortsetzung nächste Seite

Welcher der abgebildeten Vögel ist der Würger?
Kreuze den Würger an!

Welcher Schnabel gehört dem Würger?
Kreuze den Würgerschnabel an!

Welche Kopfzeichnung hat der Würger?
Kreuze den Würgerkopf an! (Damit die Aufgabe nicht zu einfach wird, sieht man die Schnäbel nicht.)

Fortsetzung nächste Seite

Das Frankenstein-Spiel

In einer erfundenen Geschichte hat Dr. Frankenstein aus Teilen von toten Menschen einen neuen Menschen zusammengenäht. In dieser Horrorstory haben die Teile nicht richtig zusammengepasst und so wurde dieser Mensch ein Monster! – Deine Aufgabe ist es nun, aus den Einzelteilen einen Würger zusammenzubauen. Suche die richtigen Vogelteile für einen Würger zusammen, so dass diesmal das Experiment gelingt und ein normaler Würger herauskommt. Kreuze alle Würgerteile an!

Aufgabe:

Fortsetzung nächste Seite

Einleitung:

Kreuze den Eisvogel an!

„Sandgruben sind so leer, so trostlos, so tot, da kann man höchstens Bundeswehrsoldaten Schießübungen machen lassen oder sie einfach fluten.“
Ist das so?

Aufgabe:


Finde mit Hilfe der Karte zum Schaufenster den Eisvogel!
Welche der Gefiederzeichnungen gehört dem Eisvogel? Kreuze die Eisvogelzeichnung an! (Damit die Aufgabe nicht zu einfach wird, sieht man die Schnäbel nicht.)

Fortsetzung nächste Seite

Welche Körpergestalt hat der Eisvogel?
Kreuze den Eisvogelkörper an!




Welcher Schnabel gehört dem Eisvogel?
Kreuze den Eisvogelschnabel an!







Fortsetzung nächste Seite










Das Frankenstein-Spiel

In einer erfundenen Geschichte hat Dr. Frankenstein aus Teilen von toten Menschen einen neuen Menschen zusammengenäht. In dieser Horrorgeschichte haben die Teile nicht richtig zusammengepasst und so wurde dieser Mensch ein Monster! – Deine Aufgabe ist es nun, aus den Einzelteilen einen Eisvogel zusammenzubauen. Suche die richtigen Vogelteile für einen Eisvogel zusammen, so dass diesmal das Experiment gelingt und ein normaler Eisvogel herauskommt. Kreuze die Eisvogelteile an!









Einleitung: Das Paradies des Jägersmann

Mit einem Zielfernrohr und einem Gewehr, das wirklich ganz genau geradeaus schießt, kann man auf der abgebildeten Wiese (Foto über den Rebhühnern) auf 500m Entfernung einem Hasentier präzise zwischen den Löffeln (Ohren) durchschießen. Nur: Auf dieser Wiese können keine Hasen leben...

Aufgabe:

Welches Hasentier siehst Du im Schaufenster? Schau' Dir kurz die Texte durch, finde den richtigen Namen des Tieres heraus und schreibe ihn auf! Feldhas


Auf dem linken Bild (Foto über den Rebhühnern) siehst Du eine Landschaft, wie man sie in Deutschland inzwischen fast überall finden kann. Auf dem mittleren Bild (Foto einer Böschung) siehst Du einen Teil einer Landschaft, der selten geworden ist.

Fortsetzung nächste Seite





Diese Landschaftselemente verbessern die Überlebenschancen von „Hasentieren“: Streiche Falsches durch!

Hecken
~~Riesenfelder~~
 Bekämpfung von Weinbergschnecken
 Abbau der Krötenschutzzäune
 Häufiges Mähen/Häufige Mahd von Wiesen
 Feldrandstreifen
 Anbringen von Nistkästen
 Felder, auf denen nur eine Pflanzensorte angebaut wird
 Große Anzahl unterschiedlicher Wildkräuter
 Besserer Fledermausschutz
 Böschungen
 Anlage von Mooren
 Mäuerchen aus Natursteinen zwischen den Feldern
 Einsatz von Großmaschinen
 Abschalten der Wasserkläranlagen
 Ungemähte kleine Flächen zwischen Feldern
 Benutzung von Schädlingsbekämpfungsmitteln
 Büsche und kleine Bäume zwischen den Feldern



Einleitung:

Eule: „Hi, ich bin Bubo, der Uhu!“
 Echse: „...und ich bin Lacerta, die Eidechse!“
 Eule: „Ich bin ein Vogel so wie eine Taube oder ein Specht.“
 Echse: „...und ich bin ein Kriechtier so wie ein Krokodil oder eine Schlange.“

Prof. Haarspalter: Der Archaeopteryx ist mit Euch beiden verwandt, weil ihr beide Körpermerkmale mit ihm gemeinsam habt.

Eule: „Das ist doch Blödsinn. Kriechtiere sind doof!“
 Echse: „Wir sind gar nicht doof, aber Schmarrn ist es trotzdem. Die hochnäsigen Vögel und wir können doch keine gemeinsamen Verwandten haben!“

Fortsetzung nächste Seite



Aufgabe:

Prof. Haarspalter hat Recht, auch wenn Bubo und Lacerta das nicht hören wollen.

Streiche bei den folgenden Aussagen immer die falsche Tiergruppe durch!

Die Zähne des Archaeopteryx weisen auf eine Verwandtschaft mit dem ☒ Vogel ☐ Reptil hin.

Die Federn des Archaeopteryx weisen auf eine Verwandtschaft mit dem ☐ Vogel ☒ Reptil hin.

Die Krallen an den Flügeln des Archaeopteryx weisen auf eine Verwandtschaft mit dem ☒ Vogel ☐ Reptil hin.

Fortsetzung nächste Seite

Hier sind Vogel- und Reptilienmerkmale dargestellt. Kreuze an, welche von den Merkmalen der Archaeopteryx hatte! (Es kann mehr als ein Merkmal sein.)

Federn ☒

Zähne ☒

Krallen an den Flügeln ☒

Schnabel ☐

Kurzes Schwänzchen ☐

...übrigens: Aus einem Ei schlüpfen sie alle!

Einleitung:

Prof. Haarspalter: „Wissenschaftler wissen was!
 Ah... In diesem Schaufenster z.B. versuchen sie ...äh... zu erklären wie fett, wie dick, äh... wie schwer Vögel werden oder ...äh... wie viele Hektar groß oder ...äh... dass man gelbe Vögel immer ...äh... in der Mitte ankleben muss oder...? Ähhhhh... Ich glaube, ich bin ein bisschen durcheinander...“
 Kannst Du Prof. Haarspalter helfen?

Aufgabe:
 Streiche jeweils die beiden falschen Antworten durch!

Wie viel Platz braucht ein Wespenbussard?
 etwa 7g ~~ca. 1 ha~~ über 1000 ha

Wie schwer ist ein Kernbeißer?
 etwa 7g knapp über 50g ~~ca. 70g~~

Fortsetzung nächste Seite

Ein Spatz ist etwa so groß wie ein Grünfink.
 Also wiegt er ~~mehr als 50g~~ weniger als 7g und die 25g.

Ein Spatz braucht wahrscheinlich eine Reviergröße von
 100ha ~~10ha~~ 1ha.


Wie viel Platz braucht ein Grünfink?
 ca. 1 ha ~~etwa 4ha~~ über 12ha

Wie schwer ist ein Wespenbussard?
 etwa 0,6kg ~~knapp über 50g~~ ca. 70g


Wie viel Platz braucht ein Kernbeißer?
 etwa 7ha ~~ca. 1 ha~~ über 1000 ha

Wie schwer ist ein Grünfink?
 etwa 25g ~~knapp über 50g~~ ca. 70g

b) SF-Treatment:




Einleitung:
Das soll ein Urwald sein?! Manche Stellen im Fichtelgebirge sehen heute auch nicht anders aus.
...außerdem: Wo sind Lianen und Riesenbäume, Papageien und Schlangen, Jaguare und Indianer mit Giftpfeilen?



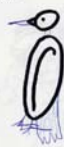
Und trotzdem: Auch im Fichtelgebirge gab es Urwald!

Aufgabe:
Der Buntspecht ist ein typischer Specht. Finde mit Hilfe der Karte zum Schaufenster den Buntspecht!
Wie viele Spechte sind in (und vor) dem Schaufenster zu sehen? 2 Stück
Welche Spechte sind in (und vor) dem Schaufenster zu sehen?
Schreibe die Antwort auf die Rückseite!

Fortsetzung nächste Seite



Wie unterscheiden sich Buchfink und Buntspecht?
Korrigiere das Vogelbild und verändere es so, dass aus dieser Strichzeichnung ein Specht wird. Achte besonders auf die Form von Schnabel, Schwanz und Fuß!



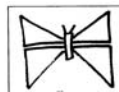
der Buntspitz ist viel kleiner und hat einen kleineren Schnabel.





Einleitung:
„Ein Klabautermann versteckt dauernd meine Kappe, Hein Blöd hat das Schiff auf eine Sandbank gesetzt und die drei Bärchen singen seit 17 Stunden... –Ich brauch' Urlaub!
„Wacholderheide im Frankenjura“??? – Klingt ja komisch! Was soll das sein?! Wo gibt's das denn?“
Wenn Du Goßweinstein, Waischenfeld oder Pottenstein kennst, dann kannst Du Blaubär weiterhelfen. Du warst bestimmt schon einmal da! Das sind alles Orte mitten in der Fränkischen Schweiz! Und genau dort, in der Fränkischen Schweiz nämlich, findet man diese „Landschaftsform „Wacholderheide“ im Frankenjura“ überall.

Aufgabe:
Der Neuntöter ist ein typischer Würger. Der Steinschmätzer ist kein Würger! Finde mit Hilfe der Karte zum Schaufenster den Neuntöter!
Wie viele Würger sind im Schaufenster zu sehen? 3 Stück
Schreibe die Namen der Würger auf die Rückseite!

Fortsetzung nächste Seite





Vergleiche Steinschmätzer und Neuntöter. Achte besonders auf die Schnäbel! Schreibe die Antwort auf die Rückseite!
Korrigiere das Vogelbild und verändere es so, dass aus dieser Strichzeichnung ein Würger wird. Achte besonders auf die Form von Schnabel und Schwanzfedern und auf die Kopfzeichnung!




Beim Neuntöter ist der Schnabel dicker und das Gefieder ist bräunlich.
Beim Steinschmätzer ist der Schnabel dünner und das Gefieder ist grünlich.

Einleitung:

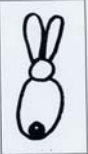



„Sandgruben sind so leer, so trostlos, so tot, da kann man höchstens Bundeswehrsoldaten Schießübungen machen lassen oder sie einfach fluten.“
Ist das so?

Aufgabe:
Finde mit Hilfe der Karte zum Schaufenster den Eisvogel!
Wie viele Eisvögel findest Du? 2 Eisvögel
Beschreibe Unterschiede zwischen Eisvogel und Bienenfresser!
Schreibe die Antwort auf die Rückseite!
Korrigiere das Vogelbild und verändere es so, dass aus dieser verunglückten Skizze ein Eisvogel wird. Achte besonders auf die Form von Schnabel und Schwanz und auf die Gefiederzeichnung des Eisvogels!

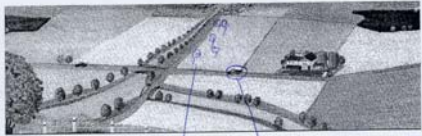


Einleitung: Das Paradies des Jägersmann

Mit einem Zielfernrohr und einem Gewehr, das wirklich ganz genau geradeaus schießt, kann man auf der abgebildeten Wiese (Foto über den Rebhühnern) auf 500m Entfernung einem Hasentier präzise zwischen den Löffeln (Ohren) durchschießen.
Nur: Auf dieser Wiese können keine Hasen leben...

Aufgabe:
Welche Tiere siehst Du im Schaufenster? Finde die Namen der Tiere heraus und notiere sie auf der Rückseite!
Überlege Dir, was diese Tiere, z.B. das Hasentier, zum Leben brauchen. Schau Dir noch einmal das Foto über den Rebhühnern an. Welche Schwierigkeiten könnten die Tiere auf diesem Feld haben? Erfahrung da viele Autos und es gibt nicht mehr viele Bäume
Dies ist ein kleiner Ausschnitt aus einer Landschaft in Deutschland.
Verändere das Bild so, dass sich Tiere dort wohl fühlen. Beschrifte Deine Veränderungen!



Aufgabe:




Einleitung:





Eule: „Hi, ich bin Bubo, der Uhu!“
Echse: „...und ich bin Lacerta, die Eidechse!“
Eule: „Ich bin ein Vogel so wie eine Taube oder ein Specht.“
Echse: „...und ich bin ein Kriechtier so wie ein Krokodil oder eine Schlange.“

Prof. Haarspalter: Der Archaeopteryx, ist mit Euch beiden verwandt, weil ihr beide Körpermerkmale mit ihm gemeinsam habt.

Eule: „Das ist doch Blödsinn. Kriechtiere sind doof!“
Echse: „Wir sind gar nicht doof, aber Schmarr ist es trotzdem. Die hochnäsigen Vögel und wir können doch keine gemeinsamen Verwandten haben!“

Fortsetzung nächste Seite

Aufgabe:





Prof. Haarspalter hat Recht, auch wenn Bubo und Lacerta das nicht hören wollen.
Schreibe auf, welche Merkmale der Archaeopteryx mit dem Fledervieh (Bubo) und welche mit dem bezahnten Echsentier (Lacerta) gemeinsam hat.

Sequenz: 2. Teil: 1. frei
Fledervieh: vorhanden
Flyer: frei
Mittelfußhaken: frei
Archaeopteryx: 2. Teil: 1. frei
Flügel: mit Knochen
Zähne: vorhanden
Kleinhorn: klein

Der Archaeopteryx ist ein Vogel

Einleitung:






Prof. Haarspalter: „Wissenschaftler wissen was!
 Äh... In diesem Schaufenster z.B. versuchen sie ...äh... zu erklären wie fett, wie dick, äh... wie schwer Vögel werden oder ...äh... wie viele Hektar groß oder ...äh... dass man gelbe Vögel immer ...äh... in der Mitte ankleben muss oder...? Ähhhhh... Ich glaube, ich bin ein bisschen durcheinander...“
 Kannst Du Prof. Haarspalter helfen?


Aufgabe:
 Vergleiche Platzbedarf und Gewicht von Pirol und Wespenbussard! Schreibe Deine Ergebnisse auf die Rückseite!
 Ein Zeisig ist ungefähr so groß wie eine Blaumeise. Wie schwer ist er wohl? Wie viel Platz braucht er ungefähr? *11g, 800 km²*
 Schau Dir die Vögel an, die hier dargestellt sind! Schätze doch einmal, wie schwer eine Taube ist? Wie viel Platz braucht sie?
140 km 150g

c) S-Treatment:

Einleitung:




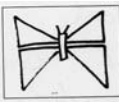
Das soll ein Urwald sein?! Manche Stellen im Fichtelgebirge sehen heute auch nicht anders aus.
 ...außerdem: Wo sind Lianen und Riesenbäume, Papageien und Schlangen, Jaguare und Indianer mit Giftfeilen?



Und trotzdem: Auch im Fichtelgebirge gab es Urwald!

Aufgabe:
 Erkläre dem Indianer den „Urwald des Fichtelgebirges“!
 Schreibe einfach auf, was Dich besonders interessiert, was vielleicht auch für Dich neu ist. Schau es Dir genau an und mache Dir Notizen. Was willst Du ihm erzählen?



Einleitung:

„Ein Klabauteermann versteckt dauernd meine Kappe, Hein Blöd hat das Schiff auf eine Sandbank gesetzt und die drei Bärchen singen seit 17 Stunden... –Ich brauch' Urlaub!
 ‚Wacholderheide im Frankenjura‘??? – Klingt ja komisch! Was soll das sein?! Wo gib't's das denn?“
 Wenn Du Gößweinstein, Waischenfeld oder Pottenstein kennst, dann kannst Du Blaubär weiterhelfen. Du warst bestimmt schon einmal da! Das sind alles Orte mitten in der Fränkischen Schweiz! Und genau dort, in der Fränkischen Schweiz nämlich, findet man diese Landschaftsform ‚Wacholderheide im Frankenjura‘ überall.

Aufgabe:
 Kannst Du Blaubär etwas zur ‚Wacholderheide im Frankenjura‘ erzählen? Schreibe auf, was auch Dich besonders interessieren würde!

Einleitung:



„Sandgruben sind so leer, so trostlos, so tot, da kann man höchstens Bundeswehrsoldaten Schießübungen machen lassen oder sie einfach fluten.“
Ist das so?

Aufgabe:

Stell Dir vor, Du wirst Dich gleich mit einem General unterhalten, der darüber entscheidet, ob das nächste Panzermanöver in der Sandgrube durchgeführt wird oder nicht. Du sollst ihm nicht sagen, dass er das Manöver nicht abhalten soll. –Darauf würde er nicht hören. Beschreibe ihm einfach, was Du schön, interessant oder aus anderen Gründen gut und wichtig findest.

Schreibe auf, was Dich besonders interessiert, was vielleicht sogar für Dich neu ist. Schau das Schaufenster genau an und mache Dir Notizen. Was erzählst Du dem General?

Einleitung: Das Paradies des Jägersmann

Mit einem Zielfernrohr und einem Gewehr, das wirklich ganz genau geradeaus schießt, kann man auf der abgebildeten Wiese (Foto über den Rebhühnern) auf 500m Entfernung einem Hasentier präzise zwischen den Löffeln (Ohren) durchschießen.




Nur: Auf dieser Wiese können keine Hasen leben...

Aufgabe:

Kannst Du dem enttäuschten Jäger erklären, wie man unsere Kulturlandschaft verändern müsste, damit Hasen und Rebhühner und viele andere Tiere hier wieder besser leben können?

Schreibe Deine Erklärung auf!

Einleitung:

Eule: „Hi, ich bin Bubo, der Uhu!“
Echse: „...und ich bin Lacerta, die Eidechse!“
Eule: „Ich bin ein Vogel so wie eine Taube oder ein Specht.“
Echse: „...und ich bin ein Kriechtier so wie ein Krokodil oder eine Schlange.“



Prof. Haarspalter: Der Archaeopteryx, ist mit Euch beiden verwandt, weil ihr beide Körpermerkmale mit ihm gemeinsam habt.

Eule: „Das ist doch Blödsinn. Kriechtiere sind doof!“
Echse: „Wir sind gar nicht doof, aber Schmarr ist es trotzdem. Die hochnäsigen Vögel und wir können doch keine gemeinsamen Verwandten haben!“

Aufgabe:

Prof. Haarspalter hat Recht, auch wenn Bubo und Lacerta das nicht hören wollen.
Kannst Du die beiden überzeugen? Schreibe Deine Argumente auf!

Einleitung:

Prof. Haarspalter: „Wissenschaftler wissen was!
Äh... In diesem Schaufenster z.B. versuchen sie ...äh... zu erklären wie fett, wie dick, äh... wie schwer Vögel werden oder ...äh... wie viele Hektar groß oder ...äh... dass man gelbe Vögel immer ...äh... in der Mitte ankleben muss oder...? Ähhhhh... Ich glaube, ich bin ein bisschen durcheinander...“

Kannst Du Prof. Haarspalter helfen?

Aufgabe:

Verstehst Du das Schaufenster?
Schau es Dir genau an! Kannst Du Prof. Haarspalter den Inhalt des Schaufensters erklären? Schreibe möglichst genau auf, was im Schaufenster dargestellt sein soll!

Anhang 2: Formale Analyse der Treatments; vergleichbare Implikationen: „≈“ ; ungleiche Implikationen: „≠“. (Einzelheiten siehe Text).

	Treatment	F		SF		S
Biotope:						
Aufgaben	Urwald	Kreuze an!	≈	Benenne! Unterscheide! Korrigiere Bild!	≠	Erkläre (...)!
Zielsetzung		konvergent		konvergent		divergent
Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Auswahlantworten: 1 aus 3 und 4 aus 12)		Einfachantworten, Mehrfachantworten		freie Gestaltung
<i>Schülerhandeln</i>		<i>Merkmale vergleichen</i>		<i>Merkmale vergleichen</i>		<i>unbestimmbar</i>
Aufgaben	Wacholderheide	Kreuze an!	≈	Benenne! Vergleiche! Korrigiere Bild!	≠	Beschreibe (...)!
Zielsetzung		konvergent		konvergent		divergent
Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Auswahlantworten: 1 aus 3 und 3 aus 9)		Einfachantworten, Mehrfachantworten		freie Gestaltung
<i>Schülerhandeln</i>		<i>Merkmale vergleichen</i>		<i>Merkmale vergleichen</i>		<i>unbestimmbar</i>
Aufgaben	Sandgrube	Kreuze an! (Multiple choice mit 1 aus 3 und Mehrfachwahl)	≈	Nenne! Beschreibe Unterschiede! Korrigiere Bild!	≠	Erzähle zu (...)!
Zielsetzung		konvergent		konvergent		divergent
Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Auswahlantworten: 1 aus 3 und 4 aus 12)		Einfachantworten, Mehrfachantworten		freie Gestaltung
<i>Schülerhandeln</i>		<i>Merkmale vergleichen</i>		<i>Merkmale vergleichen</i>		<i>unbestimmbar</i>
Themen:						
Aufgaben	Archaeopteryx	Streiche Falsches! Kreuze an!	≈	Suche Gemeinsamkeiten!	≈	Überzeuge (...)!
Zielsetzung		konvergent		konvergent		konvergent
Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Alternativantworten, Auswahlantworten: 3 aus 5)		Kombination zwischen Mehrfach- und Reihenantwort		freie Gestaltung
<i>Schülerhandeln</i>		<i>Merkmale vergleichen</i>		<i>Merkmale vergleichen</i>		<i>Merkmale vergleichen</i>
Aufgaben	Kulturlandschaft	Streiche Falsches!	≠	Überlege und nenne Schwierigkeiten! Verändere Bild!	≈	Erkläre (...)!
Zielsetzung		konvergent		divergent		divergent
Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Auswahlantworten: 11 aus 18)		freie Gestaltung		freie Gestaltung
<i>Schülerhandeln</i>		<i>rezipierte Informationen überprüfen / durchdenken</i>		<i>kreative Eigentätigkeit</i>		<i>kreative Eigentätigkeit</i>
Aufgaben	Platzbedarf	Streiche Falsches!	≈ / ≠	Vergleiche! Wie viel Platz, wie schwer?	≠	Erkläre (...)!
Zielsetzung		konvergent		konvergent		konvergent

Beantwortung		gebunden		frei		frei
Weitere Kategorisierung		Multiple Choice (Auswahlantworten: 2 aus 3)		Einfachantworten, Mehrfachantworten		freie Gestaltung
<i>Schülerhandeln</i>		<i>Lesen von Graphik</i>		<i>Lesen von Graphik, Transfer auf unbehandeltes Objekt</i>		<i>Erfassung und Vermittlung von Hauptinhalt</i>

Die obige Tabelle bezieht sich neben der Charakterisierung der in den unterschiedlichen Treatments gestellten Aufgaben auf deren methodische Implikationen bei der Umsetzung im Handeln der Schüler, also auf die Frage, worauf die Aufgabe und deren Beantwortung im Einzelnen abzielen, z.B. auf welche fachgemäßen Arbeitsweisen (Killermann 1995, S. 196 ff.), und damit auf eine weitere Analyse der drei zu unterscheidenden Treatments.

Zwischen Treatment F und SF sind beim Diorama „Urwald im Fichtelgebirge“ gleiche oder ähnliche methodische Implikationen festzustellen: Das Treatment F ist charakterisiert durch konvergente Zielsetzung, gebundene Beantwortung und Multiple-Choice-Aufgaben, das Treatment SF ebenfalls durch konvergente Zielsetzung, jedoch freie Beantwortung und Aufgaben mit Einfach- oder Mehrfachantworten. Die sich bei beiden Gruppen dahinter verbergenden fachgemäßen Arbeitsweisen sind sehr ähnlich, nämlich v. a. das Vergleichen von Merkmalen, im einen Fall ausgehend von einem bestimmten Objekt (Stopfpräparat), damit nur einer Art (Schwarzspecht), und im anderen Fall ausgehend von mehreren Präparaten, damit mehreren Arten einer Tiergruppe (Spechte). Wahrscheinlich unterschiedliche methodische Implikationen findet man beispielsweise zwischen SF und S bei „Urwald im Fichtelgebirge“: Das Treatment SF ist bestimmt durch konvergente Zielsetzung, freie Beantwortung und Aufgaben mit Einfach- oder Mehrfachantworten. Das Treatment S durch divergente Zielsetzung, freie Beantwortung und freie Gestaltung der Bearbeitung charakterisiert. Das Schülerhandeln im S-Treatment ist kaum vorherzusagen und wird selbst in dem begrenzten musealen Rahmen durch das vorgegebene Diorama kaum dem Schülerhandeln des SF-Treatments entsprechen.

Die Charakterisierung der Aufgaben kann Hinweise auf Ähnlichkeiten oder Unterschiede der methodischen Implikationen zwischen den Treatments geben, wie etwa bei der Station Kulturfolger. Zwischen Treatment F und SF finden sich keine Ähnlichkeiten, zwischen Treatment SF und S sind alle Kategorien identisch. Korrespondierend mit diesem Befund werden die methodischen Implikationen zwischen Treatment F und SF als sehr unterschiedlich, die zwischen Treatment SF und S als sehr ähnlich bewertet. Es ist jedoch genauso möglich, dass, wie dies bei allen Biotopen der Fall ist, die Aufgabencharakterisierung keine Unterschiede zwischen F/SF und SF/S zeigt, die methodischen Implikationen dennoch verschieden sind (vgl. „Urwald im Fichtelgebirge“, s.o.). Die Entscheidung, ob bezüglich methodischer Implikation Ähnlichkeit oder Unähnlichkeit angenommen wird, beruht v. a. auf dem erwarteten Schülerhandeln. Zu ergänzen ist der Vergleich von Treatment F und S. Diese sind sowohl in den Aufgaben als auch im Schülerhandeln am unähnlichsten, so dass hier, mit der Ausnahme der Station „Archaeopteryx“, von Unähnlichkeit ausgegangen werden kann.

Anhang 3: Checkliste für die Betreuerinnen im Museum, um größtmögliche Standardisierung zu erreichen.

HiWi-Checkliste
Einführung der Schüler
<ol style="list-style-type: none"> 1. Betreuerin für den ganzen Vormittag 2. Selbst kurz vorstellen, Namen der Kinder erfragen → Namensschilder schreiben lassen und anbringen. (Je nach Größe der Gesamtgruppe die Anzahl der unterschiedlich farbigen Namensschildchen wählen, so dass eine zufällige Einteilung in Kleingruppen vorweggenommen ist.) 3. 10 min pro Schaufenster, sobald Gong → Anfangen bzw. Aufhören und zum Tisch der Betreuerin → nächstens Schaufenster 4. Gesamte Zeit am Schaufenster bleiben! 5. Schaufenster Führung 6. Austeilen der Bearbeitungsbögen und Verteilen der Schüler auf die Schaufenster (Bsp.: Gruppe mit gelbem Schild arbeitet hier!) 7. Die neuen Arbeitsblätter werden einfach auf die alten gelegt und festgeklippt. Die Reihenfolge

- ist wichtig!
8. Sobald der Gong ertönt, dürft ihr anfangen!

„Standardisierte Ermahnungen“ bei Verlassen des Schaufensters (Das ist nicht zu ernst zu nehmen!):

1. „Bitte bleibt die gesamte Zeit am Schaufenster!“ (eine Bitte)
2. Ergebnisse zeigen lassen! „Ich sehe einen Fehler! Kontrolliere noch einmal alles ganz genau!“
3. „Du bleibst an Deiner Station!“ (eine Anordnung)
4. Bescheid geben! Unterstützung holen!

Arbeitsblätter / Fragebögen:

1. Die Arbeitsblätter sollen, so wie sie von den Schülern abgegeben werden, zusammengeheftet werden! (Alle Blätter eines Schülers → eine Heftung) Damit sind die Gruppen erkennbar! Alle Arbeitsblätter einer Kleingruppe werden durch eine Manschette zusammengefasst und beschriftet: Datum, Name des HiWi, Zugang.
2. Jeder HiWi teilt seiner Gruppe den Nachtest I (Fragebogen) aus, sammelt nach 20 min den Test seiner Gruppe wieder ein, bündelt und beschriftet diesen: Datum, Name, Zugang.
3. Bitte vergessen Sie nicht nach Verabschiedung der Klasse und Aufräumen im Museum, den „HiWi-Fragebogen“ auszufüllen.

Anhang 4: Rotationsplan der Betreuerinnen: Jede der elf Betreuerinnen (A bis K) sollte jedes Treatment (S, SF, F) gleich oft begleiten. Dieses Ziel wurde in nahezu idealer Weise erfüllt. Gleichzeitig sind die Beobachtungstage von Betreuerin H und I in dieser Übersicht dargestellt. U = Untersuchungstag. B = Beobachtungstag.

	A	B	C	D	E	F	G	H	I	J	K
U1	S				SF				F		
U2		F						B	B	SF	S
U3			S		F	SF		B	B		
U4		SF	F						S		
U5	SF	S		F				B	B		
U6				SF		S	F				
U7								S	B	F	SF
U8				S			SF	F			
U9	F				SF		S				
U10			SF					B	B	S	F
U11		SF			S	F		B	B		
U12	S	F						SF			
U13	F				S				SF		
U14		S			F			SF	B		
ΣS, ΣSF, ΣF, ΣB	2,1,2	2,2,2	1,1,1	1,1,1	2,2,2	1,1,1	1,1,1	1,2,1	1,1,1	1,1,1	1,1,1
								5B	7B		

Anhang 5: Experteneinschätzung zum wahrscheinlichen Vorwissen der Probanden.

In dieser Vorerhebung wurde die Wahrscheinlichkeit, mit der ein bayerischer Gymnasiast der fünften Jahrgangsstufe aufgrund seines Vorwissens die Anforderungen von *KogMit18* meistern kann, eingeschätzt. Dies erfolgte durch bayerische Gymnasiallehrer. Die praktische Erfahrung dieser beiden Experten beläuft sich auf Unterricht in wenigstens 17 Unterstufenklassen - davon sieben fünften Klassen - für die Zeitdauer von jeweils mindestens einem halben Schuljahr. Diese unabhängige Experteneinschätzung erlaubt eine Einschätzung der Objektivität dieser Beurteilungen durch Bestimmung der Signierobjektivität (vgl. Material und Methoden 4.3.1). κ (kappa) beträgt 0,769, ein Wert, der eine hohe Urteilerkonkordanz anzeigt (Diehl & Staufenbiel 2002, S. 161 f.) und damit akzeptable Objektivität ausdrückt (Bortz & Döring 1995, S. 254).

Übersicht über die Experteneinschätzung der Beantwortungswahrscheinlichkeit von *KogMit18*. 0=wahrscheinlich nicht lösbar, 1=möglicherweise lösbar; L 1=Lehrer 1, L 2=Lehrer 2; Aufschlüsselung für die einzelnen Items des erweiterten Fragebogenteils *KogMit25*: siehe unten.

	S-Items	SF-Items	F-Items
--	---------	----------	---------

Expertenbewertung	L 1	L 2	L 1	L 2	L 1	L 2
0	6	6	5	5	4	5
1	0	0	1	1	2	1

Analyse der Beantwortungswahrscheinlichkeit durch Experteneinschätzung von *KogMit25*.
 0=wahrscheinlich nicht lösbar, 1=vielleicht lösbar; Tr.=Treatment, L 1=Lehrer 1, L 2=Lehrer 2.

Biotop	Tr.	Item	L 1	L 2
Urwald	S	Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.	0	0
	SF	Buntspechte sind kleiner als Grünspechte.	0	0
	F	Bild (Brachvogelschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	0	0
Wacholderheide	S	Bild (Bartgeier) Dies ist ein Wanderfalke.	0	0
	SF	Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Der Neuntöter gehört zu den Würgern!	0	0
	F	Bild (Würgerschnabel) Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Dies ist der Schnabel eines Raubwürgers!	0	0
Sandgrube	S	Ein Wiedehopf hat eine Haube.	0	0
	SF	Eisvögel sind Höhlenbrüter.	0	0
	F	Bild (Rabenvogelschnabel) So sieht der Schnabel eines Eisvogels aus.	0	0
Kulturfolger	S	Goldammern leben in unserer Kulturlandschaft.	0	0
	SF	Autoabgase bewirken einen Rückgang von Rebhühnern.	1	1
	F	Feldhasen können ohne Hecken und hohes Gras in der freien Wildbahn überleben.	1	1
Archaeopteryx	S	Im Flügelskelett des Urvogels Archaeopteryx sind deutlich drei freie Fingerknochen zu sehen.	0	0
	SF	Schaut man sich das Knochengerüst von heute lebenden Vögeln und Kriechtieren an, so erkennt man bei beiden einen langen knöchernen Schwanz.	0	0
	F	Der Urvogel Archaeopteryx hat am Ende seiner Flügel Krallen.	1	0
Platzbedarf	S	Je kleiner der Vogel, desto größer der Platzbedarf.	0	0
	SF	Ein Zaunkönig, der kleinste Vogel Deutschlands, braucht so viel Platz wie eine Wacholderdrossel.	0	0
	F	Ein Grünfink braucht mehr Platz als ein Kernbeißer.	0	0
Urwald (Zusatz)	F	Bild (Buntspechtschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	0	0
Kindermuseum		Bild (Fischotter) Dies hier ist ein Dachs!	1	1
		Ein Schwarzstorch hat einen weißen Bauch.	0	0
		Ein Rotfuchs hat Krallen wie ein Hund.	1	0
„Carry over“		Hin und wieder sieht man in Oberfranken Takahes.	0	0
		Keas sind im Fichtelgebirge selten geworden.	0	0
		Kakapos gehören zu den flugunfähigen Vögeln.	0	0

Wie Übersicht und differenzierte Analyse bestätigen, ist die Einschätzung der beiden Experten sehr ähnlich. Schüler der fünften Klasse können diese Items aufgrund ihres Vorwissens wahrscheinlich zum größten Teil nicht beantworten.

Anhang 6: Pretest-Effekt.

Dargestellt sind deskriptive Statistiken zu Messzeitpunkten und den drei Antwortmöglichkeiten zu den Items aus dem Fragebogenteil *KogMit18*, daneben die entsprechenden Ergebnisse verteilungsfreier Teststatistik. Durchgeführt wurde der Friedman-Test. In keinem Fall ergeben sich signifikante Unterschiede oder auch nur Tendenzen. Damit ist der Pretest-Effekt zu vernachlässigen. N=53.

Mess-zeitpunkt	Lösung	Mittelwert	Standard-abweichung	Friedman-Test
Vortest	Richtig	4,70	2,72	p=,226
Nachtest I	Richtig	5,06	2,48	
Nachtest II	Richtig	5,11	3,17	
Vortest	Falsch	3,28	2,08	p=,292
Nachtest I	Falsch	2,87	1,97	
Nachtest II	Falsch	3,06	2,10	
Vortest	Weiß nicht	9,58	4,09	p=,264
Nachtest I	Weiß nicht	9,64	3,91	
Nachtest II	Weiß nicht	8,66	4,75	

Anhang 7: Antrag an das Kultusministerium und die entsprechende Antwort des Ministeriums zur Genehmigung der Untersuchung.

UNIVERSITÄT BAYREUTH

Fachbereich Biologie, Chemie, Geowissenschaften

Lehrstuhl Didaktik der Biologie

Prof. Dr. S. Klautke

Tel. 0921/55-2590 und 55-2591

Fax. 0921/55-2535

E-Mail: siegfried.klautke@uni-bayreuth.de

Universitätsstraße 30/NW I

95440 Bayreuth

Bayerisches Staatsministerium
Für Unterricht und Kultus
Referat VI/1

80333 München

Antrag auf Erteilung der Genehmigung für die Durchführung einer wissenschaftlichen Untersuchung an oberfränkischen Gymnasien

Sehr geehrte Damen und Herren,

der Lehrstuhl der Didaktik der Biologie der Universität Bayreuth beantragt, eine wissenschaftliche Untersuchung von unterschiedlich angeleiteten Exkursionen/ Unterrichtsgängen in einem Naturkundemuseum mit Gymnasialklassen der fünften Jahrgangsstufe durchführen zu dürfen.

Dabei werden die Gymnasiasten einen Schulvormittag im Naturkundemuseum und im Kinder-Wald-Museum des Umweltschutz-Informationszentrums Lindenhof in Bayreuth unterrichtet. Sie werden sich die Inhalte des Museums selbst erarbeiten, dabei aber drei unterschiedliche Zugänge beschreiten: Ein Treatment folgt einem konstruktivistischen Ansatz*, ein weiteres einem instruktivistischen und das dritte stellt eine Mischform zwischen beiden dar. Ziel ist es, den Erfolg dieser unterschiedlichen Methoden auf kognitiver und affektiver Ebene zu evaluieren. Das Untersuchungsdesign folgt dem anerkannten Muster aus Vortest, Treatment, Nachtest und Behaltentest.

Inhaltlich werden Aspekte von Fauna und Flora der einheimischen Natur- und Kulturlandschaft behandelt: Konkret werden die Biotope Wald, Wacholderheide, Weiher und Sandgrube, die Themen Kulturfolger, Archäopterix, Raumbedarf von Tieren und ausgewählte Aspekte des Kinder-Wald-Museums bearbeitet. Dabei werden Biotope der Region (z.B. Urwald im Fichtelgebirge, Wacholderheide der Fränkischen Schweiz) den Schülern kindgerecht nahegebracht. Sie lernen Wildtiere in ihrem Lebensraum kennen und werden sich mit umwelterzieherischen Fragestellungen unserer Kulturlandschaft auseinandersetzen. In zwei vorbereitenden Zulassungsarbeiten ergaben sich deutliche Hinweise auf die höhere Effektivität von angeleiteten Museumsbesuchen im Vergleich zu nicht angeleiteten. Wie diese Anleitung auszusehen hat, welche Ebene von Instruktion, Selbstbestimmung und Eigenständigkeit in der Problemlösung anzustreben ist, und was man auf welcher Ebene bewirken kann, ist noch ungeklärt.

In speziellen Voruntersuchungen wurden fünf Klassen den unterschiedlichen Treatments unterzogen. Daraus lässt sich abschätzen, dass etwa zwölf Schulklassen für die Hauptuntersuchung ausreichend sind. Diese soll im Zeitraum zwischen Oster- und Pfingstferien dieses Jahres durchgeführt werden. Die Schulen werden aus dem Landkreis Bayreuth bzw. Oberfranken ausgewählt. Der Transport erfolgt mit öffentlichen Verkehrsmitteln oder Charterbussen.

Der Aufwand für Schulen und Lehrer ist überschaubar: In der Schule ist für Vortest und Behaltenstest jeweils ein Fragebogen im zeitlichen Umfang von 15min auszufüllen. Ferner ist die zu untersuchende Schulklassse während der Busfahrten zum Museum bzw. zur Schule von einem Lehrer zu beaufsichtigen. Die Betreuung im Museum übernimmt der Lehrstuhl der Didaktik der Biologie. Es wird versichert, dass der laufende Unterricht durch diese Untersuchung nicht mehr als hier beschrieben tangiert werden wird.

Aufgrund von Erfahrungen aus den Voruntersuchungen wird ein Diorama in der Totale gefilmt, um die Dauer und Intensität der Beschäftigung der Schüler mit dem Lerninhalt der zu bearbeitenden Station abschätzen zu können. Die Aufnahmen werden bezüglich der obigen Punkte protokollarisch ausgewertet und sofort wieder gelöscht.

Bei allen Leistungserhebungen wird die Anonymität der Teilnehmer voll gewährleistet. Die Auswertung der Untersuchung lässt auf keinen Fall Rückschlüsse auf einzelne Schüler, Lehrer oder Schulen zu. Die Kooperation von Schulen und Eltern ist selbstverständlich Voraussetzung für diese Untersuchung.

Die übergeordnete Fragestellung dieser Untersuchung ist es, die Effektivität konstruktivistischer und instruktivistischer Herangehensweise im außerschulischen Lernort Naturkundemuseum auf kognitiver und affektiver Ebene zu klären. Die Ergebnisse sollten Aufschluss über die günstigste methodische Herangehensweise geben und einen Beitrag zur Verbesserung des Biologieunterrichts auch im Bereich der Museumspädagogik leisten.

In der Anlage übersenden wir die in der Voruntersuchung eingesetzten Fragebögen, die in der Hauptuntersuchung inhaltlich gleich bleiben werden. Wir bitten um Genehmigung dieser Untersuchung an oberfränkischen Gymnasien.

Mit freundlichen Grüßen

(Prof. Dr. Klautke)

.....(Wiss. Ass. Wilde)

* Konstruktivismus im Sinne von Mandl (Reinmann-Rothmeier, G. & Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In B. Weidenmann, A. Krapp, M. Hofer, G. L. Huber & H. Mandl (Hrsg.), Pädagogische Psychologie (S. 603-648). Weinheim: Beltz.)

Bayerisches Staatsministerium für Unterricht und Kultus

Bayerisches Staatsministerium für Unterricht und Kultus
80327 München

Universität Bayreuth - Lehrstuhl für Didaktik der
Biologie
z. Hd. Herrn Prof. Dr. S. Klautke
Universitätsstr. 30/NW I

95440 Bayreuth

Ihr Zeichen
Ihre Nachricht vom

04.02.2002

Bitte bei Antwort angeben
Unser Zeichen

VI/1 - O5106 - 6/13719

Telefon
(089) 2186

2716

München,

28.02.2002

Erhebungen an Schulen;
hier: Befragung von Schülern zum Thema "Museumsbesuch"

Sehr geehrter Herr Prof. Dr. Klautke,

das Staatsministerium genehmigt die in Ihrem Schreiben vom 04.02.2002 beantragte Befragung zum Thema „Museumsbesuch“.

Die Genehmigung ist an folgende Auflagen gebunden:

1. Die Schule und die betreffenden Fachlehrkräfte müssen mit der Durchführung der Befragung einverstanden sein. Die Mitwirkung aller Beteiligten ist freiwillig; sie werden darauf ausdrücklich hingewiesen. Vor der Durchführung der Befragung sind bei noch nicht volljährigen Schülern die Erziehungsberechtigten über Inhalt und Zweck der Untersuchung und über die freiwillige Teilnahme zu informieren.
2. Die datenschutzrechtlichen Bestimmungen sind zu beachten, insbesondere muss die Anonymität der Betroffenen gewahrt bleiben; aus der Auswertung der Untersuchung dürfen keine Rückschlüsse auf einzelne Schulen, Lehrkräfte, Eltern oder Schüler möglich sein.
3. Die Durchführung der Befragung darf nicht innerhalb der regulären Unterrichtszeit stattfinden.

Hausadresse
Salvatorstraße 2
80333 München

U-Bahn-Haltestelle
Odeonsplatz
U3, U4, U5, und U6

Telefon
(089) 2186-0

Telefax
(089) 2186-2800

e-mail
poststelle@stmukwk.bayern.de

- 2 -

4. Die Aufsichtspflicht liegt während der Befragung bei der Schule.
5. Der Fragebogen darf nicht vom vorgelegten Muster abweichen. Insbesondere wird auf das Verbot der kommerziellen Werbung in Schulen (Art. 84 Abs. 1 BayEUG) hingewiesen.

Mit freundlichen Grüßen



Dr. Städele
Ministerialrat

Anhang 8: Vortest / Fragebogenteil Soziodemographisch

Dieser Fragebogen stellt den ersten Teil der Untersuchung dar. Ihr habt 15min Zeit.

Bitte beantwortet diese erste Seite besonders sorgfältig!

Euer Name bleibt geheim. Schreibt in die Kästchen bitte nur Euer Geburtsdatum.

Tag Mon. Jahr

04 01 1991

Ich bin ein Mädchen ☒ Junge ☐

So lauten die ersten zwei Buchstaben des Vornamens meiner Mutter: HE

Warst Du schon einmal in einem Naturkundemuseum?

nein ☐ ja ☒

In welchem Museum denn?

Lindenhof, in der Stadt (Lindenberg)

Wie hat es Dir gefallen?	außerordentlich gut	ziemlich	mittelmäßig	kaum	gar nicht	Ich kann mich nicht erinnern.
	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Warst Du schon einmal am Lindenhof in Bayreuth?

ja ☒ nein ☐

Wenn ja, was hast Du dort gemacht? Meine Großmutter hat einen Besuch gemacht, wir sind zusammen in den Hof, den Teich und in den Wald gegangen und haben uns viel unterhalten.

Die Beantwortung des übrigen Fragebogens ist genauso einfach. Folgendes solltest Du beachten:

Wenn Du eine Antwort nicht weißt, sollst Du nicht raten.

Beispiel:

	In Grönland gibt es Elefanten.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
--	--------------------------------	---------------------------------	---------------------------------------	--

So einfach geht's! Viel Spass!!!

Anhang 9: Vortest / Fragebogenteil „Tieraffektivität“

Spielst Du lieber drinnen, z.B. am Computer, oder draußen, z.B. in der Natur?	drinnen <input type="checkbox"/>	draußen <input type="checkbox"/>	Ich kann mich nicht entscheiden! <input checked="" type="checkbox"/>
---	-------------------------------------	-------------------------------------	---

Empfindest Du gegenüber den folgenden Tieren Zuneigung oder Abneigung?







	große Zuneigung	eher Zuneigung	weder Zu- noch Abneigung	eher Abneigung	große Abneigung	
Maus	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Möwe	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Biber	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Storch	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Forelle	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Schwalbe	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Goldfisch	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Hase	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>

Empfindest Du gegenüber den folgenden Tieren Zuneigung oder Abneigung?

	große Zuneigung	eher Zuneigung	weder Zu- noch Abneigung	eher Abneigung	große Abneigung	
Specht	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Dachs	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Sumpfdotter-schnepfe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input checked="" type="checkbox"/>
Meerschweinchen	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Fischotter	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>
Wellensittich	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ich kenne das Tier nicht! <input type="checkbox"/>


Anhang 10: Vortest / Fragebogenteil KogO



Die folgenden Fragen sind nicht einfach! Wenn Du zu einem Thema nichts weißt, sollst Du auch nichts hinschreiben.



	Noch vor wenigen Hundert Jahren gab es im Fichtelgebirge Urwald. Wie sieht für Dich der Urwald im Fichtelgebirge aus? Schreibe in Stichpunkten auf, was Dir dazu einfällt!	Mehr Tannen als Laubbäume, kein Park, kein Wald und viele sehr, sehr alte Bäume.
	Was verbindest Du mit dem Lebensraum Sandgrube? Schreibe in Stichpunkten auf, was Dir dazu einfällt!	Meister Töchter, nicht viele Tiere, sehr, sehr viel Sand.
	Wie setzt sich für Dich die Wacholderheide in der Fränkischen Schweiz zusammen? Schreibe in Stichpunkten auf, was Dir dazu einfällt!	Eine unebene Landschaft, sehr trocken meist sonniger Hang und viele Wacholdersträucher und Tiere wie Schmetterlinge.
	Wie müsste unsere (Kultur-) Landschaft aussehen, damit Hasen hier wieder gut leben können? Schreibe in Stichpunkten auf, was Dir dazu einfällt!	Weniger Industrie, mehr Wald, mehr Wiesen, mehr ungestörte Natur.
	Wie hängen das Gewicht von Vögeln und ihr Platzbedarf zusammen? Antworte in einem Satz!	Das bewirkt, entscheidet, ob sie auf einem Ast sitzen oder auf einem dicken Ast.
	Welche Merkmale hat der Urvogel Archaeopteryx? Schreibe in Stichpunkten auf, was Dir dazu einfällt!	

Anhang 11: Vortest / Fragebogenteil KogMit25

Fragebogen:

	Eisvögel sind Höhlenbrüter.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Die Füße eines Uhus sind mit Federn bedeckt.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Hin und wieder sieht man in Oberfranken Takahes.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Ein Schwarzstorch hat einen weißen Bauch.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Ein Grünfink braucht mehr Platz als ein Kernbeißer.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Ein Rotfuchs hat Krallen wie ein Hund.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Dies hier ist ein Dachs!	stimmt <input checked="" type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Der Neuntöter gehört zu den Würgern!	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>

	Dies ist ein Wanderfalke.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Kakapos gehören zu den flugunfähigen Vögeln.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Ein Wiedehopf hat eine Haube.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Je kleiner der Vogel, desto größer der Platzbedarf.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Ein Zaunkönig, der kleinste Vogel Deutschlands, braucht so viel Platz wie eine Wacholderdrossel.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Feldhasen können ohne Hecken und hohes Gras in der freien Wildbahn überleben.	stimmt <input checked="" type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Dies ist der Schnabel eines Würgers!	stimmt <input checked="" type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Autoabgase bewirken einen Rückgang von Rebhühnern.	stimmt <input checked="" type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input type="checkbox"/>

	Buntspechte sind kleiner als Schwarzspechte.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	So sieht der Schnabel eines Spechtes aus.	stimmt <input type="checkbox"/>	stimmt nicht <input checked="" type="checkbox"/>	weiß nicht <input type="checkbox"/>
	Ein Urvogel/ Archaeopteryx hat am Ende seiner Flügel Krallen.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	So sieht der Schnabel eines Eisvogels aus.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Goldammern leben in unserer Kulturlandschaft.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Keas sind im Fichtelgebirge selten geworden.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>
	Schaut man sich das Knochengerüst von heute lebenden Vögeln und Kriechtieren an, so erkennt man bei beiden einen langen knöchernen Schwanz.	stimmt <input type="checkbox"/>	stimmt nicht <input type="checkbox"/>	weiß nicht <input checked="" type="checkbox"/>

Vielen Dank für Euere Hilfe! Wir sehen uns in einer Woche! Tschüß!

Anhang 12: Fragebogenteil AffMu

Euer Name bleibt geheim. Schreibt in die Kästchen bitte nur Euer Geburtsdatum.








Tag Mon. Jahr
04 05 1991

Ich bin ein Mädchen ☒ Junge ☐ So lauten die ersten zwei Buchstaben des Vornamens meiner Mutter: KL

Heute warst Du in einem Naturkundemuseum!








Wie hat Dir der heutige Besuch im Lindenhof gefallen?	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
---	---	---	--	---------------------------------------	---	---

So gut haben mir die folgenden Stationen gefallen:

 Urwald im Fichtelgebirge	außerordentlich gut ☺☺ <input checked="" type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Sandgrube	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Wacholderheide im Frankenjura	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Kulturfolger/ Kulturflüchter	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input checked="" type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Platzbedarf für Vögel	außerordentlich gut ☺☺ <input checked="" type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Archaeopteryx	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Kindermuseum	außerordentlich gut ☺☺ <input checked="" type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>





Anhang 13: Fragebogenteil AffMuUr

Bei der Station Urwald im Fichtelgebirge haben mir die folgenden Teile so gut gefallen:

 Die Vögel.	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Die übrigen Tiere.	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Die Pflanzen.	außerordentlich gut ☺☺ <input checked="" type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Die Pilze.	außerordentlich gut ☺☺ <input checked="" type="checkbox"/>	ziemlich ☺ <input type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Die Landschaft.	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Die Unordnung.	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
 Das Arbeitsblatt mit dem kleinen Indianer.	außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>

Anhang 14: Fragebogenteil *AffMuPla*

Bei der Station Platzbedarf für Vögel haben mir die folgenden Teile so gut gefallen:

		außerordentlich gut ☺☺ <input type="checkbox"/>	ziemlich ☺ <input checked="" type="checkbox"/>	mittelmäßig ☺ <input type="checkbox"/>	kaum ☹ <input type="checkbox"/>	gar nicht ☹☹ <input type="checkbox"/>	Ich kann mich nicht erinnern. <input type="checkbox"/>
	Die Vögel.						
	Die ‚mathematische Graphik‘.						
	Der Erklärungstext im Schau- fenster.						
	Das Arbeitsblatt mit Profes- sor Haarspalter.						

Anhang 15: Fragebogenteil *Freie Äußerung*

Möchtest Du die beiden Museen des Lindenhofs (Kinder-/ Naturkunde-
museum) wieder besuchen?

Ja ☒ mit ☐ nein ☐ weiß nicht ☐

Willst Du zu Deinem heutigen Museumsbesuch noch etwas sagen?

Es hat sehr gut gefallen und ich habe noch einiges dazu gelernt.

Tschüss und vielen Dank für Deine Hilfe!!!

Wenn Du noch Zeit hast, dann zeichne auf der Rückseite dieses Blattes, was Dir am Lindenhof am allerbesten gefallen hat.

Anhang 16: Faktorenanalyse

Kommunalitäten

	Anfänglich	Extraktion
Lindenhof gefallen	1,000	,543
Urwald	1,000	,461
Sandgrube	1,000	,401
Wachholderheide	1,000	,494
Kulturfolger	1,000	,373
Platzbedarf	1,000	,403
Archaeopteryx	1,000	,267
Kindermuseum	1,000	,113

Extraktionsmethode: Hauptkomponentenanalyse.

Erklärte Gesamtvarianz

	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
Komponente	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	3,054	38,180	38,180	3,054	38,180	38,180
2	,998	12,477	50,657			
3	,945	11,808	62,466			
4	,831	10,393	72,858			
5	,632	7,905	80,763			
6	,590	7,375	88,138			
7	,543	6,782	94,921			
8	,406	5,079	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

Komponentenmatrix

	Komponente
	1
Lindenhof gefallen	,737
Urwald	,679
Sandgrube	,633
Wachholderheide	,703
Kulturfolger	,611
Platzbedarf	,635
Archaeopteryx	,517
Kindermuseum	,337

Extraktionsmethode: Hauptkomponentenanalyse.

a 1 Komponenten extrahiert

Anhang 17: Beziehungen zwischen Museumsbesuch und Stationen innerhalb von *AffMu*.

Mit Hilfe einer Faktoren-Analyse wurde ein Faktor identifiziert, der 38% der Varianz erklärt. Entsprechend finden sich Korrelationen zwischen den getesteten Ebenen. Zwischen der Gesamtbewertung und allen Stationen finden sich positive Korrelationen (Nachtest I: $r=0,208$ bis $0,386$; Nachtest II: $r=0,290$ bis $0,471$; vgl. unten).

Wie die folgende Tabelle zeigt, finden sich positive Korrelationen zwischen dem Gefallen des gesamten Museumsbesuchs und den Einzelementen. Test: Kendall-Tau, alle Korrelationskoeffizienten (r) mindestens hoch signifikant.

	Museumsbesuch NT I (r)	Museumsbesuch NT II (r)
Urwald	0,386	0,471
Wacholderheide	0,311	0,454
Sandgrube	0,290	0,355
Kulturfolger	0,302	0,451
Archaeopteryx	0,208	0,290
Platzbedarf	0,233	0,359
Kindermuseum	0,260	0,360

Anhang 18: Beziehungen zwischen *AffMu* und *AffMuUr* und *AffMuPla*.

Es gibt Zusammenhänge zwischen Urwald und entsprechenden Einzelementen (Nachtest I: $r=0,195$ bis $0,349$ (6 Items) und $0,102$ (1 Item); Nachtest II: $r=0,314$ bis $0,420$ (6 Items) und $0,080$ (1 Item); vgl. unten) und zwischen Platzbedarf und entsprechenden Einzelementen (Nachtest I: $r=0,204$ bis $0,287$ (4 Items); Nachtest II: $r=0,273$ bis $0,395$ (4 Items); vgl. unten).

Wie die folgende Tabelle zeigt, finden sich positive Korrelationen zwischen dem Gefallen der Station „Urwald“ und Einzelementen der Station. Test: Kendall-Tau, alle Korrelationskoeffizienten (r) mindestens hoch signifikant außer bei Item „Unordnung“.

	Urwald Nachtest II (r)	Urwald Nachtest II (r)
Vögel	0,349	0,417
Andere Tiere	0,282	0,377
Pflanzen	0,319	0,420
Pilze	0,243	0,328
Landschaft	0,338	0,349
Unordnung	0,102	0,080
Arbeitsblatt	0,195	0,314

Wie die folgende Tabelle zeigt, finden sich positive Korrelationen zwischen dem Gefallen der Station „Platzbedarf“ und Einzelementen der Station. Test: Kendall-Tau, alle Korrelationskoeffizienten (r) mindestens hoch signifikant.

	Platzbedarf Nachtest II (r)	Platzbedarf Nachtest II (r)
Vögel	0,287	0,395
Grafik	0,204	0,296
Erklärungstext	0,235	0,356
Arbeitsblatt	0,262	0,273

Anhang 19: Affektive Vortest-Nachtest-Vergleiche durch Items mit Antwortvorgaben.

Erst in den Vorstudien wird deutlich, wie eine auch auf die Vortests zu beziehende affektive Größe aussehen kann. Eine methodische Optimierung, wie für die übrigen Testteile durchgeführt, war für diesen Fragebogenteil „*Tieraffektivität*“ deshalb nicht mehr möglich. Die im Test gestellte Frage lautet: „Empfindest Du gegenüber den folgenden Tieren Zuneigung oder Abneigung?“ Die fünfstufige Ratingskala reichte von „große Zuneigung“, „eher Zuneigung“ über „weder Zu- noch Abneigung“, „eher Abneigung“ bis zu „große Abneigung“, ergänzt um die Antwortmöglichkeit: „Ich kenne das Tier nicht!“ Die 14 Tiere, zu denen sich die Schüler äußern sollten, unterscheiden sich bezüglich folgender Merkmale bzw. Kategorisierungen: Tiere, die in den Museen des Lindenhofs (Naturkunde- oder Kindermuseum) vorkommen oder nicht; Säuger, Vögel oder Fische; Wild- oder Haustiere; reale oder fiktive Tiere; lebendige oder ausgestopfte Tiere.

	Tier	Naturkunde- Museumstier (M _N) / Kinder- Museumstier (M _K) / Nicht-Museumstier (N)	Säuger (S) / Vogel (V) / Fisch (F)	Wildtier (W) / Haustier (H)	Reales (R) / fiktives Tier (F)
1	Maus	M ¹⁵ , lebend	S	W	R
2	Hase	M _N , ausgestopft	S	W	R
3	Biber	N	S	W	R
4	Dachs	M _K , ausgestopft	S	W	R
5	Fischotter	N	S	W	R
6	Meerschweinchen	N	S	H	R
7	Specht	M _N , ausgestopft	V	W	R
8	Schwalbe	N	V	W	R
9	Storch	M _K , ausgestopft	V	W	R
10	Möwe	N	V	W	R
11	Wellensittich	N	V	H	R
12	Forelle	N	F	W	R
13	Goldfisch	N	F	H	R
14	Sumpfdotter-schnepfe	N	-	-	F

Dieser begrenzte Bereich des Fragebogens *Tieraffektivität* ist noch am ehesten als explorativ zu betrachten, wie die Heterogenität der nun dargelegten Überlegungen und Hypothesen zeigt, und wird erst in Folgeuntersuchungen das methodische Niveau der übrigen Testteile bieten können.

Folgende Hypothesen sollen überprüft werden:

- Museumstiere erreichen nach dem Museumsbesuch höhere affektive Bewertungen als davor. Und: Nicht-Museumstiere erreichen nach dem Museumsbesuch keine höheren affektiven Bewertungen als zuvor. Bzw.: Die affektive Bewertung der Museumstiere steigt von Vortest zu Nachtests im Vergleich mit Nicht-Museumstieren stärker an.
Diese Hypothesen sind zu testen durch den Vergleich der Schülerantworten zu folgenden Items, jeweils ein Museumstier vs. ein Nicht-Museumstier: 2 vs. 3, 4 vs. 5, 7 vs. 8, 9 vs. 10. Entscheidend ist die Konstanzhaltung möglichst aller Parameter außer der Untersuchungsvariable (Vorhandensein im Museum / Nicht-Vorhandensein im Museum). Das kann nur annähernd gelingen, da es sich um unterschiedliche Arten handelt. Tiere von Vergleichspaaren waren beides reale Wildtiere aus der gleichen Tiergruppe, die sich möglichst ähnlich ernähren und in etwa gleich bekannt und beliebt bei Kindern sind, gleichzeitig genügend unterschiedlich, so dass das Differenzierungsvermögen der Kinder nicht überfordert wird. Ein Vergleich zwischen Bunt- und Mittelspecht wäre sicher sinnlos. Ein Vergleichspaar war z.B. „Hase“, gemeint der Feldhase (*Lepus europaeus*), und Biber (*Castor fiber*), beides reale Wildtiere, beides Säugetiere ähnlicher Ordnung (Lagomorpha und Rodentia), die sich pflanzlich ernähren. Beide sind Kindern z.B. aus

¹⁵ Ein Terrarium mit Zwergmäusen steht im Bereich zwischen Naturkunde- und Kinder-Museum und lässt sich darum nicht zuordnen.

den Medien sehr geläufig, der Hase z.B. als Zeichentrickfigur und Osterhase, der Biber z.B. als Werbeträger einer Heimwerkerkette und aus der Zahnpastareklame.

- Die affektive Bewertung der „Kinder-Museumstiere“ steigt von Vortest zu Nachtests im Vergleich mit „Naturkunde-Museumstieren“ stärker an.
Hier sind Einschätzungen zu Tieren aus dem Naturkundemuseum gegen welche aus dem Kindermuseum zu testen: 2 vs. 4, 7 vs. 9. Mit den gewählten Beispielen, z. T. durch die beschränkte Auswahl der musealen Vorgaben diktiert, kann man im ersten Fall nur noch innerhalb der Gruppe der Säugetiere, im zweiten innerhalb der der Vögel bleiben.
- Die affektive Bewertung von Vögeln steigt von Vortest zu Nachtests im Vergleich mit Säugetieren stärker an.
Diese Hypothese leitet sich ab aus der viel stärkeren Präsenz von Vögeln in beiden Museumsteilen als der von Säugetieren. Beispielsweise findet sich in den Biotopen, die für den Museumsbesuch verwendet wurden, kein einziges Säugetier, aber jeweils fünf bis neun unterschiedliche Vogelarten. Vergleichen lassen sich darum die Schülerantworten zu den Säugern (zwei Museumstiere / zwei Nicht-Museumstiere) mit denen zu den Vögeln (zwei Museumstiere / zwei Nicht-Museumstiere): Item 2 bis 5 vs. 7 bis 10. Um Verfälschungen zu vermeiden, bleiben lebende Tiere und Haustiere unberücksichtigt.
- Die zu erwartende relative Beliebtheit von Haustieren im Vergleich zu Wildtieren nimmt von Vortest zu den Nachtests ab.
Zum einen lässt sich hier kontrollieren, inwieweit diese Wildtiere in der Beliebtheit Haustieren unterlegen sind, zum anderen lässt sich überprüfen, ob sich in den Nachtests der Beliebtheitsabstand verringert. Zu testen sind folgende Paarungen:
Wildtiere vs. Haustiere: 2 bis 5 vs. 6, 7 bis 10 vs. 11, 12 vs. 13.
- Die affektive Bewertung der im Museum gezeigten lebenden Tiere (Item 1) steigt von Vortest zu den Nachtests im Vergleich mit den ausgestopften Museumstieren stärker an.
Geht man davon aus, dass lebende originale Repräsentationsformen im nachgeahmten Wirklichkeitszusammenhang, hier sind es lebende Zwergmäuse in einem nachgeahmten Biotop, nach den von Uhlig (1962, S. 64 ff.) formulierten Gedanken des Echtheitsgefälles didaktisch wertvoller sein sollten als ausgestopfte originale Repräsentationsformen z. T. im nachgeahmten Wirklichkeitszusammenhang (Beispiel: Dachs in der Höhle), z. T. außerhalb des Wirklichkeitszusammenhangs (Beispiel: Feldhase in Schaukasten), dann ist zu mutmaßen, dass sich dies auch in der affektiven Einschätzung der Schüler zu den jeweiligen Unterrichtsgegenständen zeigt. Getestet werden sollen Tiere innerhalb ihrer Tiergruppe (Mammalia), die sich unterschiedlich stark vom Original im echten Wirklichkeitszusammenhang unterscheiden: 1 vs. 2, 1 vs. 4.
- Schüler erkennen ihnen unbekannte Tiere und ordnen diese als solche ein.
Ein nicht-reales / fiktives Tier wird als Kontrollvariable zur Überprüfung der Urteilsfähigkeit der Schüler in den Test aufgenommen, nämlich Item 14. Damit sollte dieses Item immer in die Kategorie „Ich kenne dieses Tier nicht!“ oder mindestens im Vergleich zu allen anderen bei Weitem am häufigsten dort eingeordnet werden.

Die Auswertung erweist sich als nicht sehr ergiebig. Die hier angedeutete sehr differenzierte Analyse kann nicht sinnvoll durchgeführt werden, v. a. weil die Schüler zu unpräzise zwischen Museumstieren und Nicht-Museumstieren differenzieren, z. T. nicht einmal zwischen realen und fiktiven Tieren. Weitere Schwierigkeiten ergeben sich aus zu hohen Vortestbewertungen (Deckeneffekt!), die kaum weitere Steigerungen erlauben.

Folgende Ergebnisse sind dennoch zu berichten: Der Vortest-Nachtest-Vergleich der emotionalen Disposition der Schüler zu den 14 im Fragebogen behandelten Tieren ergibt signifikante Verbesserungen bei zwölf (Nachtest I) bzw. zehn (Nachtest II) Items (vgl. unten). Bei vier Items (Goldfisch, Hase, Meerschweinchen, Wellensittich) gibt es zum Nachtest II keine signifikante Verbesserung. Bei diesen „Haustieren“ sind die Vortestwerte höher als bei den übrigen zehn Tieren, so dass es wahrscheinlich aufgrund von Deckeneffekten zu keiner oder verringerter Erhöhung kommt.

Positive emotionale Disposition zu Tieren im Vergleich zwischen Vortest und Nachtests. Vorgegeben war eine Skala von „große Zuneigung“ (5) bis „große Abneigung“ (1). Signifikante Werterhöhungen sind durch Fettdruck gekennzeichnet. Wilcoxon-Test.

	VT	NT I	p
Maus	3,45	4,02	,000
Möwe	3,06	3,40	,000
Biber	3,57	3,78	,000
Storch	3,37	3,67	,000
Forelle	2,94	3,19	,000
Schwalbe	3,33	3,57	,000
Goldfisch	3,75	3,88	,027
Hase	4,63	4,64	,823
Specht	3,39	3,63	,000
Dachs	3,33	3,63	,000
Sumpfdotterschnepfe	2,74	3,16	,002
Meerschweinchen	4,45	4,43	,601
Fischotter	3,33	3,42	,041
Wellensittich	4,12	4,20	,045

VT	NT II	p
3,45	3,83	,000
3,06	3,32	,000
3,57	3,72	,004
3,37	3,55	,006
2,94	3,20	,000
3,33	3,54	,000
3,75	3,86	,068
4,63	4,55	,002
3,39	3,56	,000
3,33	3,64	,000
2,74	3,17	,000
4,45	4,47	,912
3,33	3,58	,000
4,12	4,15	,460

Die Zunahme der Zuneigung ist unspezifisch. Als Blindprobe werden Tiere, die kein Exponat des Museums sind, und ein Fantasietier (Sumpfdotterschnepfe) in den Fragebogen aufgenommen. In beiden Fällen misst man nach dem Besuch signifikant erhöhte Zuneigung der Schüler. Bei der „Sumpfdotterschnepfe“ geben allerdings mehr als die Hälfte der Schüler an, das Tier nicht zu kennen. Zu keinem Zeitpunkt und bei keinem Item gibt es Unterschiede zwischen den Treatments (Kruskal-Wallis-Test, alle zu vergleichenden Paarungen $p=n.s.$). Auf eine weitere Auswertung dieser Daten wird verzichtet, da die für valide Aussagen nötige Differenzierungsfähigkeit der Schüler bezüglich der Exponate nicht gegeben ist.

Zusammenfassung zum Testteil „Tieraffektivität“: Der Museumsbesuch scheint eine positive Haltung zu Tieren zu fördern, dies jedoch unspezifisch und unabhängig von der Tatsache, ob ein Tier Exponat des Museums ist. Das Treatment spielt keine Rolle.

Anhang 20: Die sieben zusätzlichen Items: *KogMit18* + 7 Items = *KogMilt25*.

Kindermuseum		Bild (Fischotter) Dies hier ist ein Dachs!
		Ein Schwarzstorch hat einen weißen Bauch.
		Ein Rotfuchs hat Krallen wie ein Hund.
Zusatz-Item für F-Treatment	(Abbildung von Buntspechtschnabel)	Dies ist der Schnabel eines Schwarzspechts.
Carry over		Hin und wieder sieht man in Oberfranken Takahes.
		Keas sind im Fichtelgebirge selten geworden.
		Kakapos gehören zu den flugunfähigen Vögeln.

Anhang 21: Korrelationen zwischen der Biologienote der Schüler zum Halbjahr und den kognitiven Teilen des Fragebogens.

Die Ergebnisse zu *KogMit18* beziehen sich auf alle Probanden aus der Hauptuntersuchung, deren Noten bekannt waren (N=296 von 366 Probanden der Hauptuntersuchung). Die Ergebnisse zu *KogO* sind pro Diorama bzw. Schaukasten dargestellt. Sie beziehen sich auf dieselben Schüler. „VT“ und „NT“ bedeuten „Vortest“ und „Nachtest“.

	VT		NT I		NT II	
	r	p	r	p	r	p
Items mit Antwortvorgabe	,108	,064	,143	,014	,164	,005
Items ohne Antwortvorgabe (Summenscores)	,067	,250	,238	,001	,161	,005
Urwald	,057	,331	,057	,326	,062	,291
Sandgrube	,089	,125	,196	,001	,142	,015
Wacholderheide	,146	,012	,160	,006	,178	,002
Kulturfolger	-0,056	,340	,044	,451	-0,023	,698
Platzbedarf	,109	,062	,189	,001	,157	,007
Archäopteryx	,103	,077	,225	,000	,142	,014

Anhang 22: Einschätzung der kognitiven Stufe der Anforderung von *KogMit25*. Rp= Reproduktion, Ro=Reorganisation, T=Transfer, P=Problemlösung, Tr=Treatment.

Biotop	Tr.	Item	Bldgsrat
Urwald	S	Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.	Rp
	SF	Buntspechte sind kleiner als Grünspechte.	Ro
	F	Bild (Brachvogelschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	Rp
Wacholderheide	S	Bild (Bartgeier) Dies ist ein Wanderfalke.	Rp
	SF	Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Der Neuntöter gehört zu den Würgern!	Rp
	F	Bild (Würgerschnabel) Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Dies ist der Schnabel eines Raubwürgers!	Rp
Sandgrube	S	Ein Wiedehopf hat eine Haube.	Rp
	SF	Eisvögel sind Höhlenbrüter.	T
	F	Bild (Rabenvogelschnabel) So sieht der Schnabel eines Eisvogels aus.	Rp
Kulturfolger	S	Goldammern leben in unserer Kulturlandschaft.	Rp
	SF	Autoabgase bewirken einen Rückgang von Rebhühnern.	T
	F	Feldhasen können ohne Hecken und hohes Gras in der freien Wildbahn überleben.	T
Archaeopteryx	S	Im Flügelskelett des Urvogels Archaeopteryx sind deutlich drei freie Fingerknochen zu sehen.	Rp
	SF	Schaut man sich das Knochengerüst von heute lebenden Vögeln und Kriechtieren an, so erkennt man bei beiden einen langen knöchernen Schwanz.	P
	F	Der Urvogel Archaeopteryx hat am Ende seiner Flügel Krallen.	Rp
Platzbedarf	S	Je kleiner der Vogel, desto größer der Platzbedarf.	T
	SF	Ein Zaunkönig, der kleinste Vogel Deutschlands, braucht so viel Platz wie eine Wacholderdrossel.	T
	F	Ein Grünfink braucht mehr Platz als ein Kernbeißer.	T

Urwald (Zusatz)	F	Bild (Buntspechtschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	Rp
Kindermuseum		Bild (Fischotter) Dies hier ist ein Dachs!	Rp
		Ein Schwarzstorch hat einen weißen Bauch.	Rp
		Ein Rotfuchs hat Krallen wie ein Hund.	T
Carry over		Hin und wieder sieht man in Oberfranken Takahes.	Rp
		Keas sind im Fichtelgebirge selten geworden.	Rp
		Kakapos gehören zu den flugunfähigen Vögeln.	Rp

Anhang 23: Gewichtung der Werte bezogen auf die Bloom'sche Taxonomie.

Gewichtung	Faktor 1	Faktor 2	Faktor 3
„knowledge dimension“	terminology detail	classification specific detail generalization	principle
„cognitive process dimension“	recognizing	recalling classifying	comparing inferring executing critiquing

Anhang 24: Inhaltliche Analyse von *KogMit25* mit Hilfe der Bloom'schen Taxonomie (Anderson et al. 2001, S. 27 ff.) und zwei Versuche zur Bewertung des kognitiven Anforderungsniveaus. Weitere Erklärungen finden sich im Text. Dargestellt sind alle Items dieses Teils, die doppelt gerahmt sind von der Bewertungsroutine für die drei Treatmentgruppen auszuschließen. In der Spalte „Station“ sind die sechs Arbeitsstationen im Museum verkürzt aufgeführt. Die Kürzel in der Spalte „Matrix“, z.B. F1a, beziehen sich auf die bei der Erläuterung der Bloom'schen Taxonomie eingeführten Abkürzungen und Gliederungspunkte. Tr.=Treatment, B1=Beurteilungswert 1, B2=Beurteilungswert 2.

Station	Tr.	Items / Einzelanforderungen	B1		B2	Matrix
Urwald	S	Waldschnepfen sind braun gefleckt und haben einen langen dünnen Schnabel.	Σ2		Σ6	

		Artbezeichnung wieder erkennen können Merkmale (Zeichnung, Schnabel) der Waldschnepfe vorstellen können ¹⁶		terminology recognizing specific detail recalling	1 1 2 2	F1a F1b
	SF	Buntspechte sind kleiner als Grünspechte.	Σ5		Σ13	
		Artbezeichnungen wieder erkennen können (2x) diese Vögel in der richtigen Größe vorstellen können (2x) Größe vergleichen können		terminology recognizing detail recalling details comparing	1x2 1x2 1x2 2x2 - 3	F1a F1b F2f
	F	Bild (Brachvogelschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	Σ2		Σ5	
		Artbezeichnung wieder erkennen können Schnabel eines Schwarzspechtes wieder erkennen können		terminology recognizing specific detail recognizing	1 1 2 1	F1a F1a
Wacholderheide	S	Bild (Bartgeier) Dies ist ein Wanderfalke.	Σ2		Σ4	
		Artbezeichnung wieder erkennen können		terminology recognizing	1 1	F1a

¹⁶ Das Verb „vorstellen“ ist hier verwendet im Sinne von „sich etwas bildhaft ins Gedächtnis (zurück)rufen“, nicht im Sinne von „sich etwas ausdenken, träumen, phantasieren“.

		Habitus eines Wanderfalken wieder erkennen können		detail recognizing	1 1	F1a
	SF	Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Der Neuntöter gehört zu den Würgern!	$\Sigma 3$		$\Sigma 7$	
		Bezeichnung der Tiergruppe wieder erkennen können		terminology recognizing	1 1	F1a
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Zuordnung wieder erkennen können		classification recognizing	2 1	C1a
	F	Bild (Würgerschnabel) Es gibt eine Gruppe von Vögeln, die man als Würger bezeichnet. Dies ist der Schnabel eines Raubwürgers!	$\Sigma 2$		$\Sigma 5$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Schnabel eines Würgers wieder erkennen können		specific detail recognizing	1 2 1	F1a
Sandgrube	S	Ein Wiedehopf hat eine Haube.	$\Sigma 2$		$\Sigma 6$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Merkmal (Kopfgefieder) des Wiedehopf vorstellen können		specific detail recalling	1 2 2	F1b
	SF	Eisvögel sind Höhlenbrüter.	$\Sigma 3$		$\Sigma 11$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Detail aus dem Diorama vorstellen können		specific detail recalling	1 2 2	F1b
		Wahrscheinliche Schlussfolgerung mit Hilfe gegebener Kategorisierung durchführen können		classification inferring	2 3	C2e
	F	Bild (Rabenvogelschnabel) So sieht der Schnabel eines Eisvogels aus.	$\Sigma 2$		$\Sigma 5$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Schnabel eines Eisvogels wieder erkennen können		specific detail recognizing	1 2 1	F1a
Kulturfolger	S	Goldammern leben in unserer Kulturlandschaft.	$\Sigma 2$		$\Sigma 6$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Detail aus dem Diorama vorstellen können		specific detail recalling	1 2 2	F1b
	SF	Autoabgase bewirken einen Rückgang von Rebhühnern.	$\Sigma 5$		$\Sigma 18$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Merkmale geeigneten Lebensraums (z. B. Foto im Diorama) vorstellen können		specific detail recalling	1 2 2	F1b
		Eigenschaft (Schadwirkung von Autoabgasen) erinnern können		detail recalling	1	F1b
		Einordnung eines durch Autoabgase belasteten Lebensraumes als schlecht geeignet		categories	2	C2c
		generalisierende Schlussfolgerung durchführen können		classifying	2	
				generalization executing	2 3	C2e
	F	Feldhasen können ohne Hecken und hohes Gras in der freien Wildbahn überleben.	$\Sigma 4$		$\Sigma 15$	
		Artbezeichnung wieder erkennen können		terminology recognizing	1	F1a
		Merkmale geeigneten Lebensraums (z. B.		recognizing	1	

		Foto im Diorama) vorstellen können Einordnung eines durch Fehlen wichtiger Elemente gekennzeichneten Lebensraums als ungeeignet wahrscheinliche generalisierende Schlussfolgerung durchführen können		specific detail recalling categories classifying generalization executing	2 2 2 3	F1b C2c C2e
Archaeopteryx	S	Im Flügelskelett des Urvogels Archaeopteryx sind deutlich drei freie Fingerknochen zu sehen.	$\Sigma 2$		$\Sigma 6$	
		Gattungsbezeichnung wieder erkennen können Merkmal (freie Fingerknochen) vorstellen können		terminology recognizing specific detail recalling	1 1 2 2	F1a F1b
	SF	Schaut man sich das Knochengerüst von heute lebenden Vögeln und Kriechtieren an, so erkennt man bei beiden einen langen knöchernen Schwanz.	$\Sigma 3$		$\Sigma 11$	
		Vogelskelette vorstellen können Kriechtierskelette vorstellen können Merkmale vergleichen können		element recalling element recalling elements comparing	2 2 2 3	F1b F1b F2f
	F	Der Urvogel Archaeopteryx hat am Ende seiner Flügel Krallen.	$\Sigma 2$		$\Sigma 6$	
		Gattungsbezeichnung wieder erkennen können Merkmal (Krallen an Flügelenden) vorstellen können		terminology recognizing specific detail recalling	1 1 2 2	F1a F1b
Platzbedarf	S	Je kleiner der Vogel, desto größer der Platzbedarf.	$\Sigma 2$		$\Sigma 8$	
		Zusammenhang vorstellen können Vorstellung mit gegebener Abstraktion vergleichen können		principle recalling principle comparing	3 2 - 3	C1b C2f
	SF	Ein Zaunkönig, der kleinste Vogel Deutschlands, braucht so viel Platz wie eine Wacholderdrossel.	$\Sigma 2$		$\Sigma 8$	
		Zusammenhang vorstellen können Zusammenhang auf neue Information anwenden können		principle recalling principle executing	3 2 - 3	C1b C3a
	F	Ein Grünfink braucht mehr Platz als ein Kernbeißer.	$\Sigma 6$		$\Sigma 18$	
		Artbezeichnungen wieder erkennen können (2x) diese Vögel in der richtigen Größe vorstellen können (2x) Zusammenhang vorstellen können Zusammenhang auf unbearbeitete Vogelpaarung anwenden können		terminology recognizing detail recalling principle recalling principle executing	1x2 1x2 1x2 2x2 3 2 - 3	F1a F1b C1b C3a
Urwald (Zusatz)	F	Bild (Buntspechtschnabel) So sieht der Schnabel eines Schwarzspechtes aus.	$\Sigma 2$		$\Sigma 5$	
		Artbezeichnung wieder erkennen können Schnabel eines Schwarzspechtes wieder erkennen können		terminology recognizing specific detail recognizing	1 1 2 1	F1a F1a

Kindermus.		Bild (Fischotter) Dies hier ist ein Dachs!	$\Sigma 2$		$\Sigma 4$	
		Artbezeichnung wieder erkennen können		terminology	1	F1a
		Habitus eines Dachses wieder erkennen können		recognizing	1	
				detail	1	F1a
				recognizing	1	
		Ein Schwarzstorch hat einen weißen Bauch.	$\Sigma 2$		$\Sigma 6$	
		Artbezeichnung wieder erkennen können		terminology	1	F1a
		Merkmal (Färbung der Bauchseite) vorstellen können		recognizing	1	
				specific detail	2	F1b
				recalling	2	
		Ein Rotfuchs hat Krallen wie ein Hund.	$\Sigma 4$		$\Sigma 11$	
		Artbezeichnung wieder erkennen können		terminology	1	F1a
		Merkmal (Krallen) vorstellen können (bei Hund und Rotfuchs) (2x)		recognizing	1	
		Merkmale vergleichen können		specific detail	2	F1b
				recalling	2x2	
				details	-	F2f
				comparing	3	
Carry over		Hin und wieder sieht man in Oberfranken Takahas.	$\Sigma 1$		$\Sigma 2$	
		Sicherheit Artbezeichnung nicht wieder erkennen zu können		terminology	1	F1a
				recognizing	1	
		Keas sind im Fichtelgebirge selten geworden.	$\Sigma 1$		$\Sigma 2$	
		Sicherheit Artbezeichnung nicht wieder erkennen zu können		terminology	1	F1a
				recognizing	1	
		Kakapos gehören zu den flugunfähigen Vögeln.	$\Sigma 1$		$\Sigma 2$	
		Sicherheit Artbezeichnung nicht wieder erkennen zu können		terminology	1	F1a
				recognizing	1	
Für alle		Richtigkeit des Items beurteilen können	je 1	critiquing (deciding/ judging)	je 3	F/C 5b

Anhang 25: Korrigierte Trennschärfekoeffizienten (TSK; max.=1.0, d.h. ideale Trennschärfe) und Schwierigkeitsindices (SI) in Prozent von *KogMit18*. Mit Ausnahme von Item 3 im Vortest und Item 8 im Nachtest I sind alle Trennschärfekoeffizienten signifikant.

Vortest			Nachtest I			Nachtest II		
Item	TSK	SI	Item	TSK	SI	Item	TSK	SI
1	0,251	22,7	15	0,318	56,8	14	0,338	65,0
2	0,320	11,0	7	0,202	27,7	21	0,290	34,6
3	0,092	26,2	11	0,251	64,6	7	0,280	54,1
4	0,354	17,3	4	0,150	47,9	19	0,181	18,7
5	0,313	22,1	23	0,270	53,5	16	0,288	55,0
6	0,146	16,0	1	0,189	22,4	20	0,253	16,9
7	0,318	19,8	10	0,227	24,5	8	0,243	30,3
8	0,174	58,7	14	0,223	85,4	25	0,295	70,3
9	0,315	29,5	5	0,121	45,3	11	0,240	30,5
10	0,170	67,7	8	0,094	75,7	18	0,213	72,0
11	0,284	21,7	6	0,158	49,2	1	0,250	48,5
12	0,310	39,7	21	0,145	62,6	9	0,238	63,9
13	0,199	35,6	25	0,185	45,0	15	0,281	32,4
14	0,260	59,8	18	0,254	54,3	6	0,319	45,6
15	0,300	51,3	22	0,198	70,6	4	0,194	75,8
16	0,292	32,5	2	0,240	61,5	17	0,405	38,8
17	0,318	18,2	3	0,279	37,2	5	0,309	27,6
18	0,171	29,1	9	0,235	30,3	2	0,247	36,6

Anhang 26: Ratekorrektur nach Bortz & Döring (1995, S. 196 ff.):

Folgende korrigierte Werte ergeben sich für die Mittelwerte. Die originalen Werte stehen in eckigen Klammern [...]. Berechnet sind Lernerfolg und Treatmentunterschiede der Gesamtgruppe (N=366).

Zum Lernerfolg:

Vortest:	1,88 ± 3,31 [5,65]
Nachtest I:	5,63 ± 3,68 [9,01]
Nachtest II:	4,43 ± 3,92 [8,07]

Friedmann-Test: $p < 0,001$ [.001]

Wilcoxon-Test: $p(\text{VT-NT I}) < 0,001$ [.001]

$p(\text{VT-NT II}) < 0,001$ [.001]

$p(\text{NT I-NT II}) < 0,001$ [.001]

Zu Treatmentunterschieden:

Kruskal-Wallis-Test:

Vortest:	$p = 0,329$ [.796]
Nachtest I:	$p < 0,001$ [.001];

Mann-Whitney-U-Test: $p(\text{S/SF}) < 0,001$ [.001], $p(\text{S/F}) = 0,047$ [.005], $p(\text{SF/F}) = 0,024$ [.021].

Nachtest II:	$p = 0,001$ [.002];
--------------	---------------------

Mann-Whitney-U-Test: $p(\text{S/SF}) < 0,001$ [.001], $p(\text{S/SF}) = 0,577$ [.406], $p(\text{SF/F}) = 0,003$ [.008].

Weder beim Lernerfolg noch bei den Treatmentunterschieden finden sich qualitative Unterschiede. Es wird keine Korrektur der Ratewahrscheinlichkeit vorgenommen.

Anhang 27:**KogMit18 / KogO:****Gesamtgruppe**

Kolmogorov-Smirnov-Anpassungstest

		Richtig VT	Richtig NT	Richtig BT	Summen-score VT	Summen-score NT	Summen-score BT
N		366	366	366	366	366	366
Parameter der Normalverteilung	Mittelwert	5,65	9,01	8,07	5,2432	10,6503	9,3443
	Standardabweichung	3,035	2,999	3,397	3,73922	5,73937	6,06689
Extremste Differenzen	Absolut	,134	,086	,085	,125	,057	,072
	Positiv	,134	,063	,085	,125	,057	,072
	Negativ	-,065	-,086	-,076	-,080	-,032	-,062
Kolmogorov-Smirnov-Z		2,557	1,652	1,623	2,386	1,090	1,372
Asymptotische Signifikanz (2-seitig)		,000	,009	,010	,000	,186	,046

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Lindenhof-Kenner

Kolmogorov-Smirnov-Anpassungstest

		Richtig VT	Richtig NT	Richtig BT	Summen-score VT	Summen-score NT	Summen-score BT
N		121	121	121	121	121	121
Parameter der Normalverteilung	Mittelwert	6,17	9,54	9,00	6,1240	11,2231	9,5289
	Standard-	3,249	3,031	3,445	4,36381	5,68549	5,80097

	ab- weichung						
Extremste Differ- enzen	Absolut	,111	,082	,085	,148	,098	,074
	Positiv	,111	,063	,085	,148	,098	,074
	Negativ	-,096	-,082	-,075	-,080	-,049	-,050
Kolmogorov- Smirnov-Z		1,223	,902	,938	1,625	1,082	,809
Asymptotische Signifikanz (2- seitig)		,101	,389	,342	,010	,192	,530

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Lindenhof-Nicht-Kenner

Kolmogorov-Smirnov-Anpassungstest

		Richtig VT	Richtig NT	Richtig BT	Summen- score VT	Summen- score NT	Summen- score BT
N		241	241	241	241	241	241
Parameter der Normalverteilung	Mittelwert	5,36	8,72	7,59	4,7801	10,3029	9,2448
	Standard- ab- weichung	2,892	2,944	3,290	3,25406	5,74851	6,22179
Extremste Differ- enzen	Absolut	,143	,098	,089	,106	,065	,074
	Positiv	,143	,062	,085	,106	,065	,074
	Negativ	-,064	-,098	-,089	-,071	-,037	-,069
Kolmogorov- Smirnov-Z		2,220	1,527	1,383	1,648	1,010	1,150
Asymptotische Signifikanz (2- seitig)		,000	,019	,044	,009	,260	,142

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

AffMu:

Gesamtgruppe

Kolmogorov-Smirnov-Anpassungstest

Nachtest I:

		Lindenhof- gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx
N		348	357	359	356	353	358	361
Parameter der Normalverteilung	Mittelwert	4,45	3,84	3,88	3,78	3,93	3,90	3,93
	Standard- ab- weichung	,740	,856	,853	,923	,965	,971	1,068
Extremste Differ- enzen	Absolut	,347	,248	,267	,208	,238	,203	,220
	Positiv	,230	,198	,209	,187	,150	,161	,157
	Negativ	-,347	-,248	-,267	-,208	-,238	-,203	-,220
Kolmogorov- Smirnov-Z		6,476	4,678	5,067	3,921	4,471	3,846	4,171
Asymptotische Signifikanz (2- seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Nachtest II:

		Linden- hof- gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx
N		348	354	358	335	347	351	346
Parameter der Normalverteilung	Mittelwert	4,14	3,71	3,66	3,55	3,76	3,76	3,90
	Standard- ab- weichung	,880	,845	,970	1,065	1,007	,982	1,053
Extremste Differ- enzen	Absolut	,263	,299	,241	,217	,235	,245	,230
	Positiv	,180	,224	,169	,144	,154	,166	,147
	Negativ	-,263	-,299	-,241	-,217	-,235	-,245	-,230
Kolmogorov- Smirnov-Z		4,902	5,622	4,566	3,975	4,386	4,585	4,281
Asymptotische Signifikanz (2- seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Linden-
hof-Kenner:

Kolmogorov-Smirnov-Anpassungstest

Nachtest I:

		Linden- hof- gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx
N		118	115	117	116	116	118	118
Parameter der Normalverteilung	Mittelwert	4,36	3,74	3,83	3,62	3,91	3,97	3,84
	Standard- ab- weichung	,822	,899	,864	,948	1,009	,995	1,021
Extremste Differ- enzen	Absolut	,309	,249	,271	,201	,267	,215	,198
	Positiv	,217	,195	,208	,201	,156	,149	,159
	Negativ	-,309	-,249	-,271	-,199	-,267	-,215	-,198
Kolmogorov- Smirnov-Z		3,354	2,669	2,928	2,161	2,873	2,335	2,154
Asymptotische Signifikanz (2- seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Nachtest II:

		Linden- hof- gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx
N		117	117	118	108	114	117	113
Parameter der Normalverteilung	Mittelwert	4,04	3,77	3,66	3,65	3,82	3,75	3,90
	Standard- ab- weichung	,770	,770	,980	1,035	,983	,973	1,043
Extremste Differ- enzen	Absolut	,290	,302	,237	,235	,232	,207	,263
	Positiv	,249	,237	,170	,145	,154	,173	,146
	Negativ	-,290	-,302	-,237	-,235	-,232	-,207	-,263

Kolmogorov-Smirnov-Z		3,135	3,262	2,574	2,441	2,480	2,243	2,794
Asymptotische Signifikanz (2-seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Lindenhof-Nicht-Kenner:

Kolmogorov-Smirnov-Anpassungstest

Nachtest I:

		Lindenhof gefallen	Urwald	Sandgrube	Wachhold- erheide	Kulturfolger	Platzbe- darf	Archae- opteryx
N		227	238	238	236	233	236	239
Parameter der Normalverteilung	Mittelwert	4,50	3,88	3,89	3,86	3,94	3,86	3,98
	Standard- ab- weichung	,694	,836	,850	,905	,936	,960	1,079
Extremste Differenzen	Absolut	,367	,243	,265	,210	,225	,202	,238
	Positiv	,236	,198	,209	,181	,153	,171	,172
	Negativ	-,367	-,243	-,265	-,210	-,225	-,202	-,238
Kolmogorov-Smirnov-Z		5,530	3,746	4,095	3,222	3,437	3,103	3,681
Asymptotische Signifikanz (2-seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Nachtest II:

		Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx
N		227	233	236	223	229	230	229
Parameter der Normalverteilung	Mittelwert	4,19	3,68	3,67	3,49	3,73	3,76	3,89
	Standard- ab- weichung	,933	,887	,969	1,082	1,023	,993	1,062
Extremste Differenzen	Absolut	,252	,292	,245	,206	,236	,260	,215
	Positiv	,193	,214	,170	,150	,153	,174	,147
	Negativ	-,252	-,292	-,245	-,206	-,236	-,260	-,215
Kolmogorov-Smirnov-Z		3,799	4,461	3,769	3,083	3,569	3,949	3,255
Asymptotische Signifikanz (2-seitig)		,000	,000	,000	,000	,000	,000	,000

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Anhang 28: Schreiben an die Schüler, verlesen vor dem ersten Vortest.

Eine wissenschaftliche Untersuchung...

In einer Woche werdet ihr das Naturkundemuseum am Lindenhof besuchen. Dies ist erbeten worden von der Uni Bayreuth, die diesen Museumsbesuch untersuchen möchte.

Mein Name ist Matthias Wilde. Ich bin eigentlich Lehrer, aber im Augenblick Wissenschaftler an einer der Abteilungen des Fachbereiches Biologie an der Universität. Die Abteilung heißt ‚Didaktik der Biologie‘.

Ich werde diese Untersuchung durchführen. Bei einer wissenschaftlichen Untersuchung muss man wie bei einem Spiel bestimmte Regeln befolgen. Darum werden einige Dinge bei diesem Museumsbesuch ein bisschen anders sein als beim normalen Unterricht.

Passt gut auf, jetzt kommen die Regeln!

Spielregeln!

1. Nicht alle machen das selbe. Im Museum wird Euer Klasse in verschiedene Gruppen aufgeteilt. Diese Gruppen werden unterschiedliche Arbeitsaufträge bekommen.
2. Manche Aufgaben werden vielleicht sehr leicht sein andere sehr schwierig. Ihr sollt Euch anstrengen und alles so gut bearbeiten wie ihr könnt. Aber ihr braucht keine Angst zu haben, wenn ihr an irgendeiner Stelle nicht fertig werdet oder vielleicht in einem Ausnahmefall eine Aufgabe gar nicht bearbeiten könnt. Es gibt keine Noten!!!
3. Schreibt nicht ab! Das würde die Ergebnisse der Untersuchung unbrauchbar machen! Außerdem bringt es keinerlei Vorteil, weil ihr ja keine Noten bekommt.
4. Eine sehr wichtige Regel lautet: Nichts, was in der Untersuchung behandelt wird, darf im normalen Unterricht besprochen werden. Wundert Euch also nicht, wenn Euer Lehrer bei manchen Fragen nicht wie gewohnt anfängt zu erklären, sondern die Frage einfach nicht beantwortet.
5. Auch während der Untersuchung dürfen manche Fragen nicht beantwortet werden. Manchmal werden also ich oder meine Mitarbeiter eine Frage nicht beantworten können. (Natürlich sollt ihr trotzdem fragen, wenn ihr irgendwo Schwierigkeiten habt! Meistens werden wir Euch helfen können, ohne der Untersuchung zu schaden.)

Ich hoffe, Ihr habt Lust, mir bei der Untersuchung zu helfen und haltet Euch an die Regeln.

Herzlichen Dank schon im Voraus!

M. Wilde

Anhang 29: Exemplarische Darstellung einer Möglichkeit des Ablaufs einer Arbeitsphase, z.B. Arbeitsphase I / Treatment S. Kgr=Kleingruppe von 3 oder 4 Kindern, St=Station, W=Wechsel der Station.

Kgr					
a	St 1	W	St 2	W	St 3
b	St 2	W	St 3	W	St 1
c	St 3	W	St 1	W	St 2
Dauer	10 min		10 min		10 min

Anhang 30: Bildung der Summenscores durch *Normierung auf Maxima* im Vergleich zur Bildung der Summenscores mit den *Rohdaten*. Trotz der - absolut gesehen - unterschiedlichen Werte, sind die Aussagen identisch: 1. Schüler lernen, behalten und vergessen. 2. Es gibt jeweils Treatmentunterschiede im Nachtest I zwischen S-Schülern und den beiden übrigen Treatmentgruppen SF und F.

	VT	NT I	NT II	Friedmann	VT-NT I	VT-NT II	NT I –NT II
<i>Summenscores (Maxima)</i>	6,68	15,40	13,62	.000	.000	.000	.000
Kruskal-Wallis	.975	.003	.151				
Mann-Whitney-U		S-SF: .003 S-F: .002 SF-F: .779					
	VT	NT I	NT II	Friedmann	VT-NT I	VT-NT II	NT I –NT II
<i>Summenscores (Rohdaten)</i>	5,24	10,65	9,34	.000	.000	.000	.000
Kruskal-Wallis	.940	.001	.259				
Mann-Whitney-U		S-SF: .002 S-F: .001 SF-F: .817					

Anhang 31: Treatmentbezogene Itemkonstruktion spielt bei Lindenhof-Nicht-Kennern eine Rolle.

Zunächst zu den F-Items (vgl. Abbildung unten): Auffällig ist das besonders gute Abschneiden der F-Schüler im Nachtest I. Die Differenzen sind statistisch bedeutsam. Im Folgetest nivelliert sich der Unterschied zu den übrigen Schülern, ist zumindest nicht mehr statistisch signifikant (Vortest: Kruskal-Wallis-Test; $p=0,872$; Nachtest I: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S/SF)=0,010$, $p(S/F)<0,001$, $p(SF/F)=0,019$; Nachtest II: Kruskal-Wallis-Test; $p=0,052$).

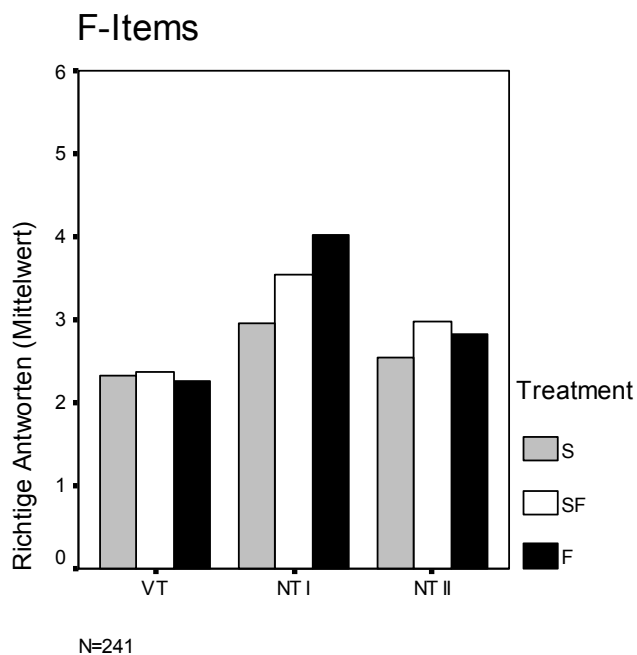


Abbildung: Lernerfolg der Lindenhof-Nicht-Kenner bezüglich der F-Items. Erläuterungen im Text. N=241.

SF-Items (vgl. Abbildung unten): Die Gruppe der SF-Schüler schneidet im Nachtest I und II signifikant besser ab als die übrigen Gruppen (Vortest: Kruskal-Wallis-Test; $p=0,840$; Nachtest I: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S/SF)<0,001$, $p(S/F)=0,755$, $p(SF/F)<0,001$; Nachtest II: Kruskal-Wallis-Test; $p<0,001$; Mann-Whitney-U-Test; $p(S/SF)<0,001$, $p(S/F)=0,184$, $p(SF/F)=0,009$).

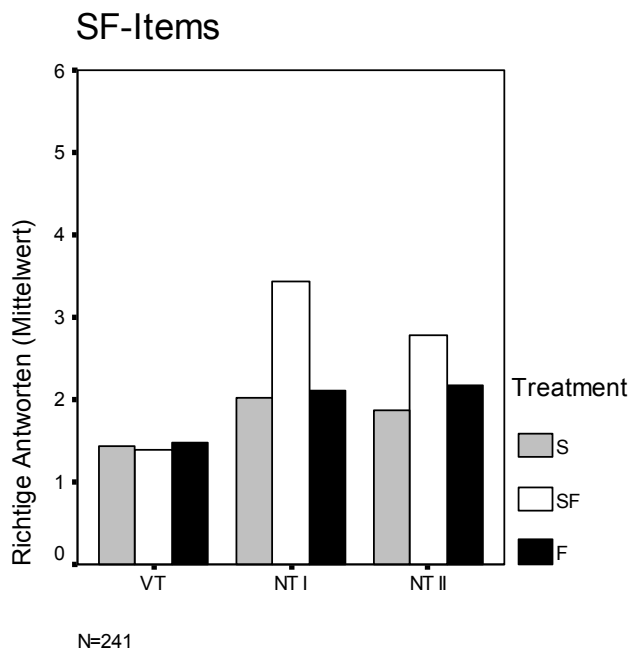


Abbildung: Lernerfolg der Lindenhof-Nicht-Kenner bezüglich der SF-Items. Erläuterungen im Text. N=241.

S-Items (vgl. Abbildung unten): Zwar erreichen die S-Schüler bei diesen Items verglichen mit F- und SF-Items die besten Ergebnisse, dennoch sind sie in Nachtest I und II in etwa gleich gut wie die F- und SF-Schüler. Keiner der Unterschiede ist statistisch abzusichern (Vortest: Kruskal-Wallis-Test; $p=0,898$, Nachtest I: Kruskal-Wallis-Test; $p=0,245$; Nachtest II: Kruskal-Wallis-Test; $p=0,132$).

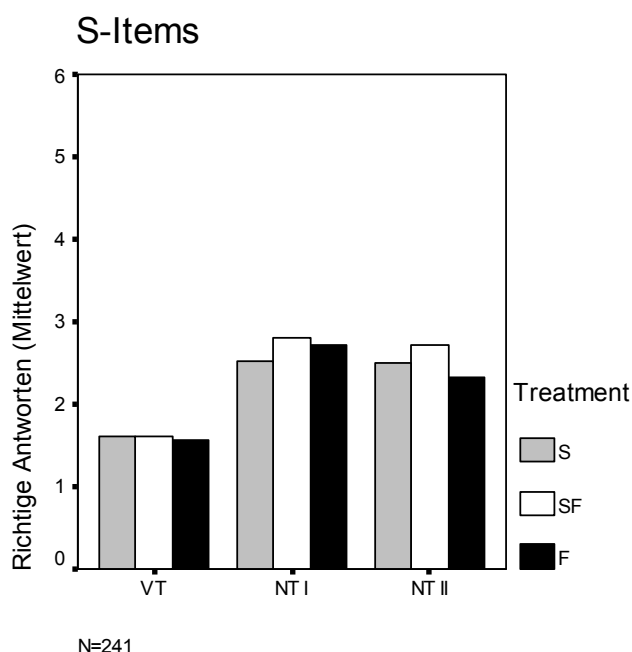


Abbildung: Lernerfolg der Lindenhof-Nicht-Kenner bezüglich der S-Items. Erläuterungen im Text.

Die Itemkonstruktion hat demnach Einfluss auf das Abschneiden der Treatmentgruppen. Die Werte in dem Wissenstest der jeweils in der Itemkonstruktion besonders beachteten Gruppe sind mindestens

genauso hoch, meist signifikant höher als die der übrigen. Von dieser treatmentbezogenen Fragebogenkonstruktion profitieren die SF-Schüler am meisten, die S-Schüler am wenigsten.

Anhang 32: Kreuztabellen zum Chi-Quadrat-Test zu Vorerfahrung der Lindenhof-Kenner („Wenn [Du den Lindenhof schon kennst], was hast Du dort gemacht?“) und durchlaufenem Treatment.

			Treatment		
			S	SF	F
Wenn bekannt, was gemacht?	keine Antwort	Anzahl	11	9	13
		Erwartete Anzahl	10,9	10,4	11,7
	Kindergeburtstag/Kinderkurs	Anzahl	9	12	11
		Erwartete Anzahl	10,6	10,0	11,4
	Biolog. Kurs	Anzahl	8	2	9
		Erwartete Anzahl	6,3	6,0	6,8
	Ausstellung	Anzahl	7	9	7
		Erwartete Anzahl	7,6	7,2	8,2
	Mehrfachbesuch	Anzahl	2	4	1
		Erwartete Anzahl	2,3	2,2	2,5
	Sonstiges	Anzahl	3	2	2
		Erwartete Anzahl	2,3	2,2	2,5
Σ		Anzahl	40	38	43
		Erwartete Anzahl	40,0	38,0	43,0

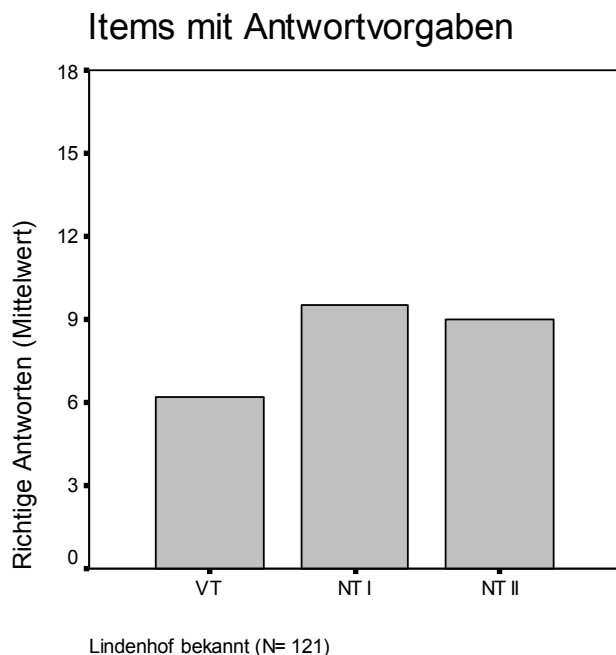
Relevant ist wegen Bedingung a die Signifikanz nach dem exakten Fisher-Test. In 58,9% der Fälle ergibt sich der Chi-Quadrat-Wert rein zufällig.

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	8,180	10	,611	,630		
Likelihood-Quotient	8,931	10	,539	,606		
Exakter Test nach Fisher	8,495			,589		
Zusammenhang linear-mit-linear	,547	1	,459	,462	,242	,022
Anzahl der gültigen Fälle	121					

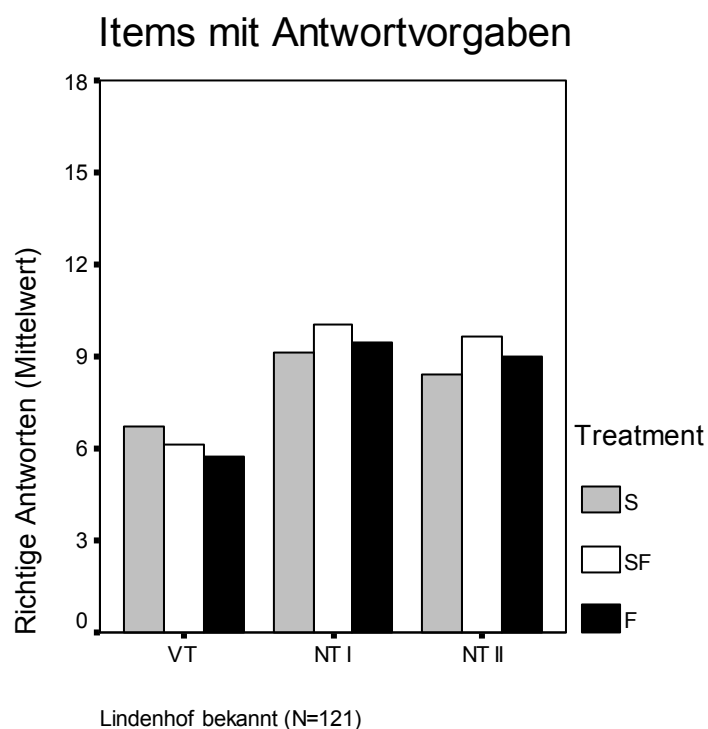
a 6 Zellen (33,3%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 2,20.

b Die standardisierte Statistik ist -,740.

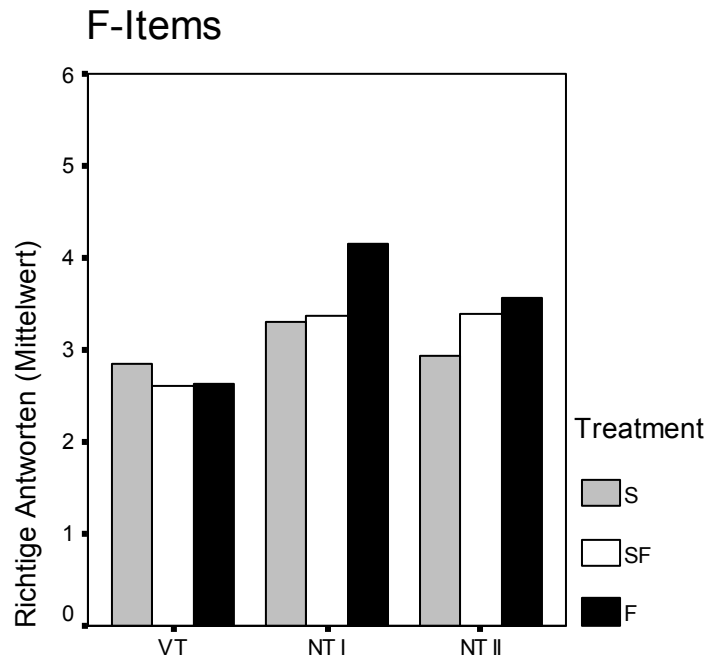
Anhang 33: Lernerfolg der Lindenhof-Kenner. (Friedman; $p < 0,001$; Wilcoxon; $p(\text{VT-NTI}) < 0,001$, $p(\text{VT-NTII}) < 0,001$, $p(\text{NTI-NTII}) = 0,055$). Weitere Erläuterungen im Text.)



Anhang 34: Vergleich zwischen den Treatmentgruppen bei den Lindenhof-Kennern (Kruskal-Wallis; Vortest, Nachtest I und II: $p = \text{n.s.}$ Weitere Erläuterungen im Text.).

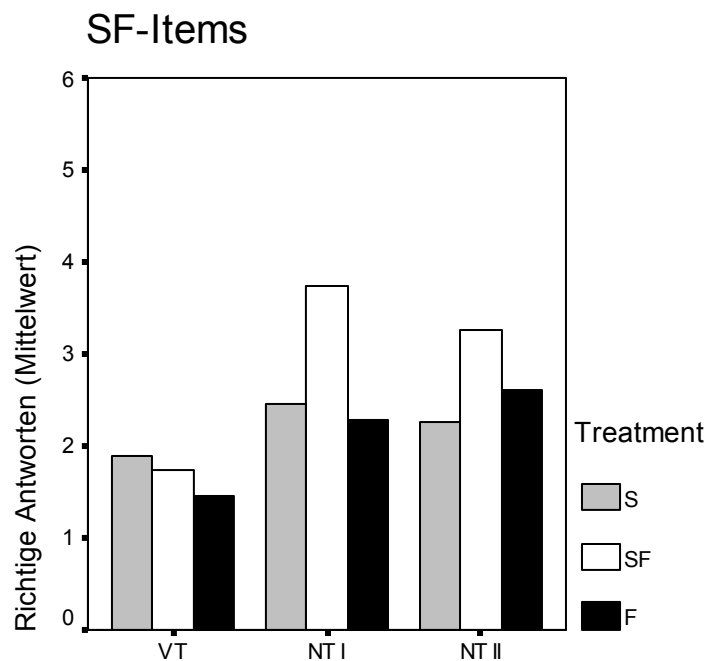


Anhang 35: Lernerfolg der Lindenhof-Kenner bezüglich der F-Items (Vortest: Kruskal-Wallis; $p=n.s.$; Nachtest I: Kruskal-Wallis; $p<0,05$; Man-Whitney-U; $p(S/SF)=n.s.$, $p(S/F)<0,05$, $p(SF/F)<0,01$; Nachtest II: Kruskal-Wallis; $p=0,106$. Weitere Erläuterungen im Text.).



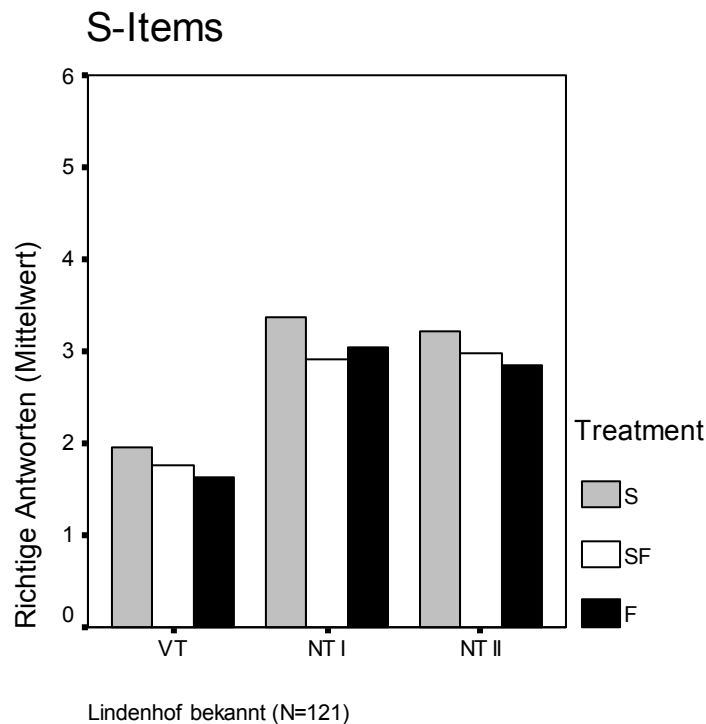
Lindenhof bekannt (N=121)

Lernerfolg der Lindenhof-Kenner bezüglich der SF-Items (Vortest: Kruskal-Wallis; $p=n.sign.$; Nachtest I: Kruskal-Wallis; $p<0,001$; Man-Whitney-U; $p(S/SF)<0,001$, $p(S/F)=n.s.$, $p(SF/F)<0,001$; Nachtest II: Kruskal-Wallis; $p<0,01$; Man-Whitney-U; $p(S/SF)<0,01$, $p(S/F)=n.s.$, $p(SF/F)<0,05$. Weitere Erläuterungen im Text.)



Lindenhof bekannt (N=121)

Lernerfolg der Lindenhof-Kenner bezüglich der S-Items (Kruskal-Wallis; Vortest, Nachtest I und II: $p=n.s.$ Weitere Erläuterungen im Text.).



Anhang 36: Gegenüberstellung der Ergebnisse der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner zu den Items mit Antwortvorgabe (*KogMit18*)

In einer Zusammenschau (vgl. Tabelle unten) zeigen sich Gemeinsamkeiten und Unterschiede zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern:

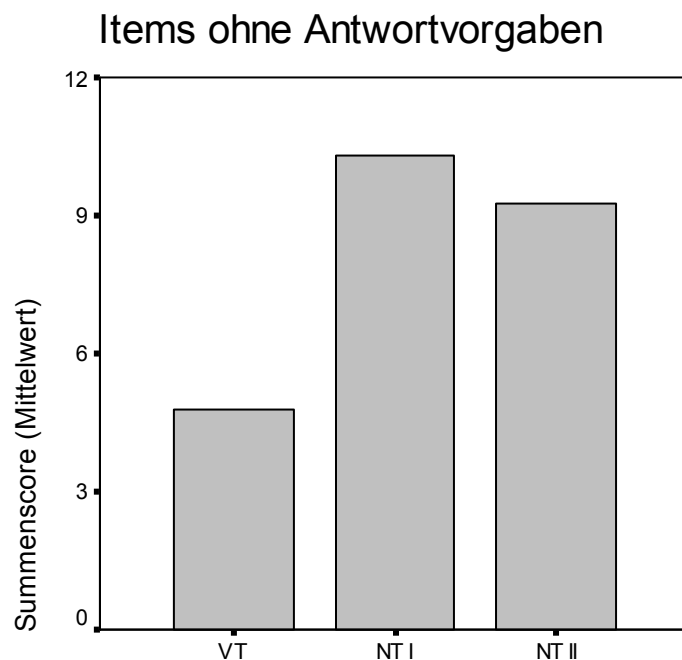
Erwerb und Persistenz von Wissen lässt sich bei beiden Gruppen belegen, Vergessen jedoch nur bei Lindenhof-Nicht-Kennern. Bei Lindenhof-Nicht-Kennern fehlt nur bei S-Schülern der Nachweis von Vergessen, bei Lindenhof-Kennern dagegen gilt dies für alle Treatmentgruppen. Folgerichtig gibt es bei Lindenhof-Kennern keine Treatmentunterschiede. Bei Lindenhof-Nicht-Kennern zeigen Schüler des Treatments SF in Nachtest I und II die besten Resultate, S-Schüler die schlechtesten. Dies ist z. T. auf die Itemkonstruktion zurückzuführen. SF-Schüler hatten Vorteile, S-Schüler Nachteile. Das gilt jedoch für Lindenhof-Nicht-Kenner wie für Lindenhof-Kenner, wo es ja keine Treatmentunterschiede gibt.

Tabelle: Übersicht über die Gegenüberstellung innerhalb der gebundenen Aufgabenbeantwortungen (Items mit Antwortvorgabe) zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern mit bilanzierender Wertung. „+“=„trifft zu“, „-“=„trifft nicht zu“, „G“=„gleich bzw. in der Kernaussage kein Unterschied“, „U“=„ungleich bzw. qualitativer Unterschied in der Aussage“. Weiter Erläuterungen im Text. N=366 oder wie angegeben.

	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz
Lernerfolg			
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	-	U
	S, N=81	S, N=40	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	-	-	G
	SF, N=81	SF, N=38	
Erwerb	+	+	G
Persistenz	+	+	G

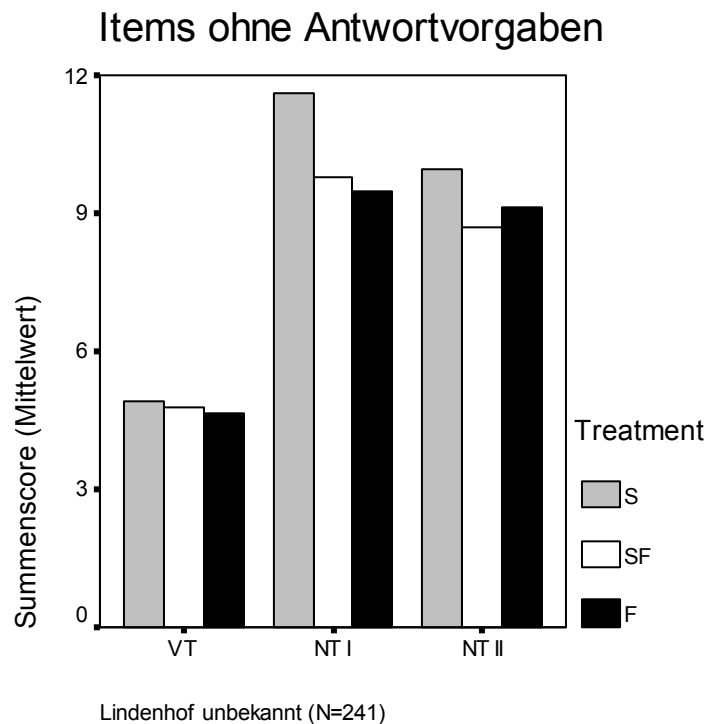
Verlust	+	-	U
	F, N=79	F, N=43	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	-	U
	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz
Treatment- unterschiede			
VT	Keine Unterschiede	Keine Unterschiede	G
NT I	SF>F>S	Keine Unterschiede	U
NT II	SF>F=S	Keine Unterschiede	U
	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	Bilanz
Fragetypenana- lyse			
S-Items			
VT	Keine Unterschiede	Keine Unterschiede	G
NT I	Keine Unterschiede	Keine Unterschiede	G
NT II	Keine Unterschiede	Keine Unterschiede	G
SF-Items			
VT	Keine Unterschiede	Keine Unterschiede	G
NT I	SF>F=S	SF>F=S	G
NT II	SF>F=S	SF>F=S	G
F-Items			
VT	Keine Unterschiede	Keine Unterschiede	G
NT I	F>SF>S	F>SF=S	Ähnlich
NT II	Keine Unterschiede	Keine Unterschiede	G

Anhang 37: Lernerfolg der Lindenhof-Nicht-Kenner bei KogO. Die Schüler gewinnen Wissen hinzu (VT<NTI), mindestens ein Teil davon ist persistent (VT<NTII) und ein Teil geht wieder verloren (NTI>NTII) (Abbildung; Friedman; $p<0,001$; Wilcoxon; alle zu vergleichenden Paare: $p<0,01$). N=241

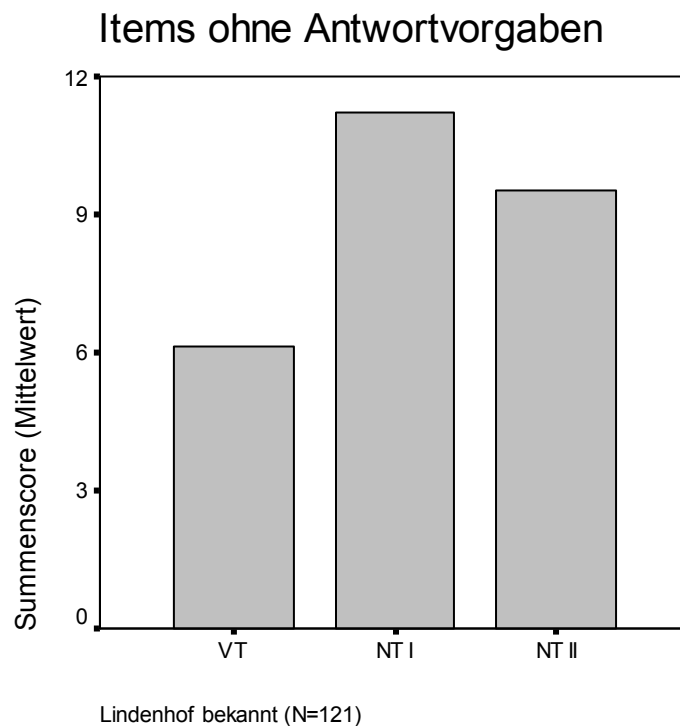


Lindenhof unbekannt (N=241)

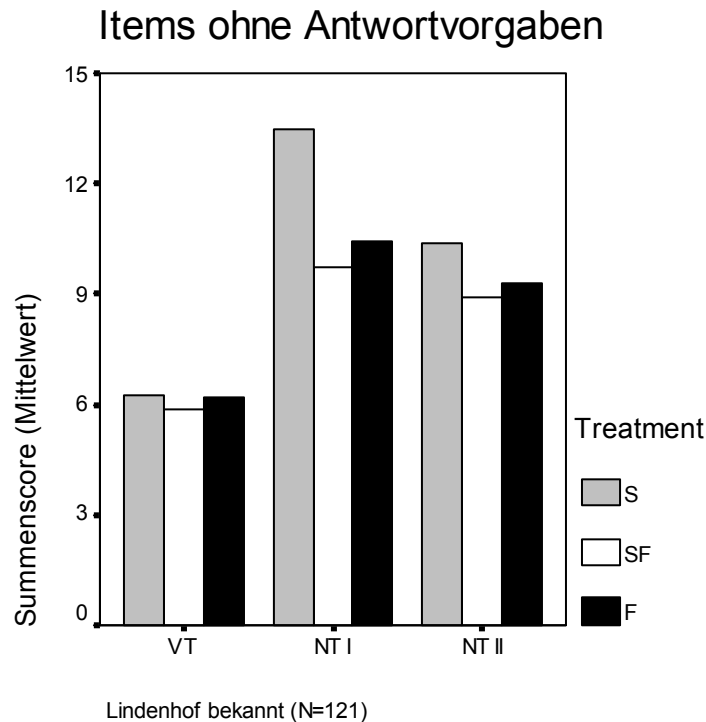
Anhang 38: Vergleich zwischen den Treatmentgruppen bei den Lindenhof-Nicht-Kennern bei *KogO*. Die drei Ausgangswerte im Vortest sind fast identisch (Kruskal-Wallis; $p=0,990$). Für Nachtest I lassen sich statistisch bedeutsame Differenzen nachweisen, für Nachtest II nicht (Kruskal-Wallis; Nachtest I: $p=0,037$, Nachtest II: $p=0,452$). Nachtest I: Mann-Whitney-U-Test: $p(S/SF)=0,044$, $p(S/F)=0,018$, $p(SF/F)=0,665$. Weitere Erläuterungen im Ergebnisteil. $N=241$



Anhang 39: Lernerfolg der Lindenhof-Kenner bei *KogO*. Die Schüler gewinnen Wissen hinzu ($VT < NTI$), ein Teil ist persistent ($VT < NTII$) und ein Teil geht wieder verloren ($NTI > NTII$) (Friedman; $p < 0,001$; Wilcoxon; alle zu vergleichenden Paare: $p < 0,001$). $N=121$.



Anhang 40: Vergleich zwischen den Treatmentgruppen bei den Lindenhof-Kennern bei *KogO*. Die Ausgangswerte unterscheiden sich nicht (Abbildung; Kruskal-Wallis; $p=0,901$). Für Nachtest I lassen sich statistisch bedeutsame Differenzen nachweisen, für Nachtest II nicht (Kruskal-Wallis; Nachtest I: $p=0,013$, Nachtest II: $p=0,414$). Nachtest I: Mann-Whitney-U-Test: $p(S/SF)=0,007$, $p(S/F)=0,016$, $p(SF/F)=0,754$; Weitere Erläuterungen im Ergebnisteil. $N=121$.



Anhang 41: Gegenüberstellung der Einzelergebnisse von Lindenhof-Nicht-Kennern und Lindenhof-Kennern zu den Items ohne Antwortvorgabe (*KogO*); vgl. Tabelle unten.

Die Verfahren zur Datenerhebung der Items ohne Antwortvorgabe liefert zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern weit weniger Unterschiede als Items mit Antwortvorgabe: Beide zeigen Erwerb, Persistenz und Verlust von Wissen. Beide weisen bei S-Schülern Vergessen, bei F-Schülern kein Vergessen auf. Nur bei SF-Schülern vergessen Lindenhof-Nicht-Kenner, Lindenhof-Kenner dagegen nicht. Beide lassen nur im Nachtest I Treatmentunterschiede erkennen, beide zu Gunsten des Treatments S.

Tabelle: Übersicht über die Gegenüberstellung innerhalb der freien Aufgabenbeantwortungen (Items ohne Antwortvorgabe) zwischen Lindenhof-Nicht-Kennern und Lindenhof-Kennern mit bilanzierender Wertung. „+“=„trifft zu“, „-“=„trifft nicht zu“, „G“=„gleich bzw. in der Kernaussage kein Unterschied“, „U“=„ungleich bzw. qualitativer Unterschied in der Aussage“, übrige Abkürzungen wie gehabt. Weitere Erläuterungen im Text. $N=366$ oder wie angegeben.

	Lindenhof-Nicht-Kenner, $N=241$	Lindenhof-Kenner, $N=121$	Bilanz
Lernerfolg			
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	+	G
	S, $N=81$	S, $N=40$	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	+	G
	SF, $N=81$	SF, $N=38$	
Erwerb	+	+	G
Persistenz	+	+	G
Verlust	+	-	U

	F, N=79	F, N=43	
Erwerb	+	+	G
Persistenz (VT<NT II)	+	+	G
Verlust	-	-	G
	Lindenhof-Nicht-Kenner, N=241	Lindenhof-Kenner, N=121	(2. Ebene)
Treatmentun- terschiede			
VT	Keine Unterschiede	Keine Unterschiede	
NT I	S>SF=F	S>SF=F	G
NT II	Keine Unterschiede	Keine Unterschiede	G

Anhang 42: Treatmentunterschiede Geschlecht; *KogMit*

Gesamtgruppe: nur Mädchen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	63	93,50
	SF	53	88,55
	F	61	84,75
	Gesamt	177	
Richtig NT	S	63	74,21
	SF	53	109,03
	F	61	86,87
	Gesamt	177	
Richtig BT	S	63	85,82
	SF	53	96,59
	F	61	85,69
	Gesamt	177	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	,923	13,590	1,679
df	2	2	2
Asymptotische Signifikanz	,630	,001	,432

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	63	48,52	3057,00
	SF	53	70,36	3729,00
	Gesamt	116		

Statistik für Test

	Richtig NT
Mann-Whitney-U	1041,000
Wilcoxon-W	3057,000
Z	-3,500
Asymptotische Signifikanz (2- seitig)	,000

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	63	57,69	3634,50
	F	61	67,47	4115,50
	Gesamt	124		

Statistik für Test

	Richtig NT
Mann-Whitney-U	1618,500
Wilcoxon-W	3634,500
Z	-1,522
Asymptotische Signifikanz (2-seitig)	,128

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	SF	53	65,67	3480,50
	F	61	50,40	3074,50
	Gesamt	114		

Statistik für Test

	Richtig NT
Mann-Whitney-U	1183,500
Wilcoxon-W	3074,500
Z	-2,479
Asymptotische Signifikanz (2-seitig)	,013

a Gruppenvariable: Treatment

Gesamtgruppe: nur Jungen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	60	94,22
	SF	67	94,72
	F	61	94,53
	Gesamt	188	
Richtig NT	S	60	74,27
	SF	67	107,38
	F	61	100,25
	Gesamt	188	
Richtig BT	S	60	78,41
	SF	67	113,10
	F	61	89,90
	Gesamt	188	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	,003	12,868	13,645
df	2	2	2
Asymptotische Signifikanz	,999	,002	,001

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	60	52,35	3141,00
	SF	67	74,43	4987,00

	Gesamt	127		
Richtig BT	S	60	51,92	3115,50
	SF	67	74,81	5012,50
	Gesamt	127		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	1311,000	1285,500
Wilcoxon-W	3141,000	3115,500
Z	-3,392	-3,512
Asymptotische Signifikanz (2-seitig)	,001	,000

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	60	52,42	3145,00
	F	61	69,44	4236,00
	Gesamt	121		
Richtig BT	S	60	56,98	3419,00
	F	61	64,95	3962,00
	Gesamt	121		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	1315,000	1589,000
Wilcoxon-W	3145,000	3419,000
Z	-2,684	-1,260
Asymptotische Signifikanz (2-seitig)	,007	,208

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	SF	67	66,95	4485,50
	F	61	61,81	3770,50
	Gesamt	128		
Richtig BT	SF	67	72,28	4843,00
	F	61	55,95	3413,00
	Gesamt	128		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	1879,500	1522,000
Wilcoxon-W	3770,500	3413,000
Z	-,788	-2,502
Asymptotische Signifikanz (2-seitig)	,431	,012

a Gruppenvariable: Treatment

Lindenhof-Kenner: nur Mädchen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	19	35,95
	SF	18	22,83
	F	21	29,38
	Gesamt	58	
Richtig NT	S	19	29,39
	SF	18	29,72
	F	21	29,40

	Gesamt	58	
Richtig BT	S	19	30,13
	SF	18	26,72
	F	21	31,31
	Gesamt	58	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	5,634	,005	,764
df	2	2	2
Asymptotische Signifikanz	,060	,998	,683

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Lindenhof-Kenner: nur Jungen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	21	30,17
	SF	20	38,65
	F	22	27,70
	Gesamt	63	
Richtig NT	S	21	27,98
	SF	20	37,10
	F	22	31,20
	Gesamt	63	
Richtig BT	S	21	26,83
	SF	20	39,90
	F	22	29,75
	Gesamt	63	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	4,096	2,632	5,784
df	2	2	2
Asymptotische Signifikanz	,129	,268	,055

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Lindenhof-Nicht-Kenner: nur Mädchen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	44	58,93
	SF	35	66,66
	F	40	55,35
	Gesamt	119	
Richtig NT	S	44	46,93
	SF	35	78,99
	F	40	57,76
	Gesamt	119	
Richtig BT	S	44	57,69
	SF	35	68,77
	F	40	54,86
	Gesamt	119	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	2,110	17,252	3,388
df	2	2	2
Asymptotische Signifikanz	,348	,000	,184

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	44	31,03	1365,50
	SF	35	51,27	1794,50
	Gesamt	79		

Statistik für Test

	Richtig NT
Mann-Whitney-U	375,500
Wilcoxon-W	1365,500
Z	-3,910
Asymptotische Signifikanz (2-seitig)	,000

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	44	38,40	1689,50
	F	40	47,01	1880,50
	Gesamt	84		

Statistik für Test

	Richtig NT
Mann-Whitney-U	699,500
Wilcoxon-W	1689,500
Z	-1,627
Asymptotische Signifikanz (2-seitig)	,104

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	SF	35	45,71	1600,00
	F	40	31,25	1250,00
	Gesamt	75		

Statistik für Test

	Richtig NT
Mann-Whitney-U	430,000
Wilcoxon-W	1250,000
Z	-2,887
Asymptotische Signifikanz (2-seitig)	,004

a Gruppenvariable: Treatment

Lindenhof-Nicht-Kenner: nur Jungen

Ränge

	Treatment	N	Mittlerer Rang
Richtig VT	S	36	62,75
	SF	46	55,41
	F	39	65,97
	Gesamt	121	
Richtig NT	S	36	43,74
	SF	46	68,54
	F	39	68,04
	Gesamt	121	
Richtig BT	S	36	48,99
	SF	46	71,97
	F	39	59,15
	Gesamt	121	

Statistik für Test

	Richtig VT	Richtig NT	Richtig BT
Chi-Quadrat	2,067	12,599	8,925
df	2	2	2
Asymptotische Signifikanz	,356	,002	,012

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	36	31,96	1150,50
	SF	46	48,97	2252,50
	Gesamt	82		
Richtig BT	S	36	33,36	1201,00
	SF	46	47,87	2202,00
	Gesamt	82		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	484,500	535,000
Wilcoxon-W	1150,500	1201,000
Z	-3,234	-2,748
Asymptotische Signifikanz (2-seitig)	,001	,006

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	S	36	30,28	1090,00
	F	39	45,13	1760,00
	Gesamt	75		
Richtig BT	S	36	34,13	1228,50
	F	39	41,58	1621,50
	Gesamt	75		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	424,000	562,500
Wilcoxon-W	1090,000	1228,500
Z	-2,969	-1,494
Asymptotische Signifikanz (2-seitig)	,003	,135

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Richtig NT	SF	46	43,08	1981,50
	F	39	42,91	1673,50
	Gesamt	85		
Richtig BT	SF	46	47,60	2189,50
	F	39	37,58	1465,50
	Gesamt	85		

Statistik für Test

	Richtig NT	Richtig BT
Mann-Whitney-U	893,500	685,500
Wilcoxon-W	1673,500	1465,500
Z	-,031	-1,877
Asymptotische Signifikanz (2-seitig)	,975	,061

a Gruppenvariable: Treatment

Anhang 43: Gefallen von Stationen und Einzelementen von Urwald und Platzbedarf: AffMuUr, AffMuPla

Nachtest I	Anzahl	Mittelwert
Urwald	366	3,84
Vögel bei Urwald	366	4,12
Tiere	366	4,09
Pflanzen	366	3,84
Pilze	366	3,39
Landschaft	366	4,39
Unordnung	366	3,52
Arbeitsblatt	366	3,48
Platzbedarf	366	3,90
Vögel bei Platzbedarf	366	4,37
Grafik	366	3,37
Erklärungstext	366	3,57
Arbeitsblatt	366	3,78

Nachtest II	Anzahl	Mittelwert
Urwald	366	3,71
Vögel bei Urwald	366	3,88
Tiere	366	4,02
Pflanzen	366	3,69
Pilze	366	3,38
Landschaft	366	4,13
Unordnung	366	3,46
Arbeitsblatt	366	3,11
Platzbedarf	366	3,76
Vögel bei Platzbedarf	366	4,04
Grafik	366	3,25
Erklärungstext	366	3,47
Arbeitsblatt	366	3,49

Anhang 44: Treatmentunterschiede im Gefallen des Gesamtbesuchs und der Stationen in der Gesamtgruppe; AffMu.

Ränge

	Treatment	N	Mittlerer Rang
Lindenhof gefallen	S	116	173,47
	SF	115	177,94

	F	117	172,14
	Gesamt	348	
Urwald	S	122	169,00
	SF	117	181,74
	F	118	186,63
	Gesamt	357	
Sandgrube	S	123	175,31
	SF	117	188,42
	F	119	176,57
	Gesamt	359	
Wachholderheide	S	123	169,21
	SF	115	176,69
	F	118	189,95
	Gesamt	356	
Kulturfolger	S	122	189,36
	SF	111	159,94
	F	120	180,22
	Gesamt	353	
Platzbedarf	S	122	175,02
	SF	117	169,38
	F	119	194,04
	Gesamt	358	
Archaeopteryx	S	124	183,97
	SF	118	175,66
	F	119	183,20
	Gesamt	361	
Kindermuseum	S	123	187,97
	SF	119	181,20
	F	122	178,25
	Gesamt	364	
Lindenhof gefallen	S	122	171,80
	SF	114	176,80
	F	112	175,10
	Gesamt	348	
Urwald	S	123	164,26
	SF	118	182,77
	F	113	186,41
	Gesamt	354	
Sandgrube	S	123	172,98
	SF	118	183,14
	F	117	182,69
	Gesamt	358	
Wachholderheide	S	116	172,44
	SF	110	171,70
	F	109	159,55
	Gesamt	335	
Kulturfolger	S	121	184,25
	SF	115	162,93
	F	111	174,30
	Gesamt	347	
Platzbedarf	S	122	171,69
	SF	116	158,88
	F	113	198,23
	Gesamt	351	
Archaeopteryx	S	117	174,99
	SF	115	183,40
	F	114	161,99
	Gesamt	346	
Kindermuseum	S	123	188,30

	SF	117	174,46
	F	116	172,19
	Gesamt	356	

Statistik für Test
Nachtest I

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeo- pteryx	Kinder- museum
Chi- Quadrat	,273	2,131	1,328	2,771	5,573	4,060	,509	1,487
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,872	,345	,515	,250	,062	,131	,775	,476

a Kruskal-Wallis-Test

b Gruppenvariable: Zugang

Nachtest II

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kinder- museum
Chi- Quadrat	,178	3,854	,829	1,337	2,935	9,926	2,928	3,353
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,915	,146	,661	,512	,230	,007	,231	,187

Anhang 45: Hauptinhalt der zeichnerischen Äußerungen, Angaben in Prozent (gerundet). N=366.

	NT I	NT II
Keine Zeichnung	20	52
Unbrauchbar	15	12
Kindermuseum	43	18
Naturkundemuseum	7	5
Beide Museen	10	5
Zwergmaus	4	8
Sonstiges	1	-

Anhang 46: Detailanalyse der zeichnerischen Äußerungen im Nachtest I. Angabe der genauen Anzahl der Nennungen geordnet nach Häufigkeiten. Doppelnennungen waren möglich. N=366.

Platzierung	Kategorie	Anzahl
1	Kletterbaum	128
2	Höhle	66
3	Vogel	60
4	Möbiliar im Kindermuseum	32
5	Sonstiges Tier	26
6	Zwergmaus	17
7	Spiel	12
	Fuchs	12
9	Kulturfolger	10
10	Stimmtelefon	6
11	Sandgrube	5
12	Archaeopteryx	3
13	Platzbedarf	2

	Lindenhof außen	2
	Pflanze	2

Anhang 47: Zusammenhang zwischen Treatment und Hauptinhalt der Zeichnung.

Hauptinhalt * Treatment Kreuztabelle

			Treatment			Gesamt
			S	SF	F	
Hauptinhalt	nichts	Anzahl	27	20	28	75
		Erwartete Anzahl	25,4	24,6	25,0	75,0
	Kindermuseum	Anzahl	58	53	46	157
		Erwartete Anzahl	53,2	51,5	52,3	157,0
	Naturkundemuseum	Anzahl	11	11	5	27
		Erwartete Anzahl	9,1	8,9	9,0	27,0
	beide	Anzahl	13	5	17	35
		Erwartete Anzahl	11,9	11,5	11,7	35,0
	nicht Lindenhof	Anzahl	10	25	19	54
		Erwartete Anzahl	18,3	17,7	18,0	54,0
	Zwergmaus	Anzahl	5	4	7	16
		Erwartete Anzahl	5,4	5,2	5,3	16,0
	nicht erkennbar	Anzahl	0	1	0	1
		Erwartete Anzahl	,3	,3	,3	1,0
	Betreuerin	Anzahl	0	1	0	1
		Erwartete Anzahl	,3	,3	,3	1,0
Gesamt		Anzahl	124	120	122	366
		Erwartete Anzahl	124,0	120,0	122,0	366,0

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	23,210	14	,057	,	
Likelihood-Quotient	24,983	14	,035	,	
Exakter Test nach Fisher	,			,	
Zusammenhang linear-mit-linear	2,411	1	,121	,	,
Anzahl der gültigen Fälle	366				

a 6 Zellen (25,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,33.

b Kann nicht berechnet werden, da zuwenig Arbeitsspeicher vorhanden ist.

Anhang 48: Zusammenhang zwischen Treatment und der Tatsache, ob gezeichnet wird.

Nachtest I:

Kreuztabelle

			Treatment			Gesamt
			S	SF	F	
Zeichnung	nein	Anzahl	27	20	28	75
		Erwartete Anzahl	25,4	24,6	25,0	75,0
	ja	Anzahl	97	100	94	291
		Erwartete Anzahl	98,6	95,4	97,0	291,0
Gesamt		Anzahl	124	120	122	366
		Erwartete Anzahl	124,0	120,0	122,0	366,0

Chi-Quadrat-Tests

	Wert	df	Asymp-totische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	1,656	2	,437	,431		
Likelihood-Quotient	1,697	2	,428	,431		
Exakter Test nach Fisher	1,682			,431		
Zusammenhang linear-mit-linear	,049	1	,824	,875	,443	,061
Anzahl der gültigen Fälle	366					

a 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 24,59.

b Die standardisierte Statistik ist -,222.

Nachtest II:

Kreuztabelle

			Treatment			Gesamt
			S	SF	F	
Zeichnung	nein	Anzahl	70	51	69	190
		Erwartete Anzahl	64,4	62,3	63,3	190,0
	ja	Anzahl	54	69	53	176
		Erwartete Anzahl	59,6	57,7	58,7	176,0
Gesamt		Anzahl	124	120	122	366
		Erwartete Anzahl	124,0	120,0	122,0	366,0

Chi-Quadrat-Tests

	Wert	df	Asymp-totische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	6,337	2	,042	,043		
Likelihood-Quotient	6,349	2	,042	,043		
Exakter Test nach Fisher	6,308			,043		
Zusammenhang linear-mit-linear	,000	1	,996	1,000	,523	,051

Anzahl der gültigen Fälle 366

a 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 57,70.

b Die standardisierte Statistik ist -,005.

Anhang 49: Gegenüberstellung der Lindenhof-Nicht-Kenner und der Lindenhof-Kenner der Items mit Antwortvorgabe im affektiven Bereich. Mann-Whitney-U-Test.

	NT I			NT II		
	Lindenhof-Nicht-Kenner	Lindenhof-Kenner	p	Lindenhof-Nicht-Kenner	Lindenhof-Kenner	p
Gesamt	4,50	4,36	,127	4,19	4,04	,017
Urwald	3,88	3,74	,218	3,68	3,77	,541
Wacholderheide	3,86	3,62	,027	3,49	3,65	,221
Sandgrube	3,89	3,83	,534	3,67	3,66	,986
Kulturfolger	3,94	3,91	,969	3,73	3,82	,533
Archaeopteryx	3,98	3,84	,130	3,89	3,90	,901
Platzbedarf	3,86	3,97	,253	3,76	3,75	,749

Anhang 50: Gegenüberstellung der Veränderungen zwischen Nachtest I und Nachtest II bei Lindenhof-Nicht-Kennern und Lindenhof-Kennern. Signifikante Veränderungen sind durch Fettdruck gekennzeichnet. Wilcoxon-Test.

	Lindenhof-Nicht-Kenner			Lindenhof-Kenner		
	NT I	NT II	p	NT I	NT II	p
Gesamt	4,50	4,19	,000	4,36	4,04	,000
Urwald	3,88	3,68	,000	3,74	3,77	,910
Wacholderheide	3,86	3,49	,000	3,62	3,65	,917
Sandgrube	3,89	3,67	,000	3,83	3,66	,123
Kulturfolger	3,94	3,73	,006	3,91	3,82	,092
Archaeopteryx	3,98	3,89	,122	3,84	3,90	,698
Platzbedarf	3,86	3,76	,210	3,97	3,75	,008

Anhang 51: Treatmentunterschiede beim Gefallen des Gesamtbesuchs und der Stationen bei Lindenhof-Nicht-Kennern. Durchgeführt wurden zuerst Kruskal-Wallis-Tests und bei signifikanten Unterschieden, d. h. $p < 0,05$, Vergleiche der Treatments mit Mann-Whitney-U-Tests

NT I:

	Lindenhofgefallen	Urwald	Sandgrube	Wachholderheide	Kulturfolger	Platzbedarf	Archaeopteryx
Chi-Quadrat	1,198	5,012	,336	1,362	2,702	5,273	,166
df	2	2	2	2	2	2	2
Asymptotische Signifikanz	,549	,082	,845	,506	,259	,072	,920

a Kruskal-Wallis-Test

b Gruppenvariable: Zugang

Keine weiteren Tests, da keine Signifikanz.

NT II:

	Lindenhofgefallen	Urwald	Sandgrube	Wachholderheide	Kulturfolger	Platzbedarf	Archaeopteryx
Chi-Quadrat	,714	7,454	2,710	2,372	,751	8,144	2,897
df	2	2	2	2	2	2	2
Asymptotische Signifikanz	,700	,024	,258	,306	,687	,017	,235

a Kruskal-Wallis-Test

b Gruppenvariable: Zugang]]

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Platzbedarf	S	80	81,15	6492,00
	SF	79	78,84	6228,00
	Gesamt	159		
Urwald	S	81	72,63	5883,00
	SF	79	88,57	6997,00
	Gesamt	160		
Platzbedarf	S	80	82,48	6598,50
	SF	77	75,38	5804,50
	Gesamt	157		

Statistik für Test

	Platzbedarf	Urwald	Platzbedarf
Mann-Whitney-U	3068,000	2562,000	2801,500
Wilcoxon-W	6228,000	5883,000	5804,500
Z	-,333	-2,315	-1,036
Asymptotische Signifikanz (2-seitig)	,739	,021	,300

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Platzbedarf	S	80	72,69	5815,00
	F	77	85,56	6588,00
	Gesamt	157		
Urwald	S	81	70,21	5687,00
	F	73	85,59	6248,00
	Gesamt	154		
Platzbedarf	S	80	71,22	5697,50
	F	73	83,34	6083,50
	Gesamt	153		

Statistik für Test

	Platzbedarf	Urwald	Platzbedarf
Mann-Whitney-U	2575,000	2366,000	2457,500
Wilcoxon-W	5815,000	5687,000	5697,500
Z	-1,860	-2,334	-1,783
Asymptotische Signifikanz (2-seitig)	,063	,020	,075

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Platzbedarf	SF	79	71,36	5637,50
	F	77	85,82	6608,50
	Gesamt	156		
Urwald	SF	79	77,77	6144,00
	F	73	75,12	5484,00
	Gesamt	152		
Platzbedarf	SF	77	66,23	5099,50
	F	73	85,28	6225,50
	Gesamt	150		

Statistik für Test

	Platzbedarf	Urwald	Platzbedarf
Mann-Whitney-U	2477,500	2783,000	2096,500

Wilcoxon-W		5637,500	5484,000	5099,500
Z		-2,092	-,411	-2,860
Asymptotische Signifi- kanz (2-seitig)		,036	,681	,004

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Urwald	S	80	74,83	5986,50
	SF	81	87,09	7054,50
	Gesamt	161		

Statistik für Test

	Urwald
Mann-Whitney-U	2746,500
Wilcoxon-W	5986,500
Z	-1,786
Asymptotische Signifi- kanz (2-seitig)	,074

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Urwald	S	80	72,13	5770,50
	F	77	86,14	6632,50
	Gesamt	157		

Statistik für Test

	Urwald
Mann-Whitney-U	2530,500
Wilcoxon-W	5770,500
Z	-2,050
Asymptotische Signifi- kanz (2-seitig)	,040

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Urwald	SF	81	78,37	6348,00
	F	77	80,69	6213,00
	Gesamt	158		

Statistik für Test

	Urwald
Mann-Whitney-U	3027,000
Wilcoxon-W	6348,000
Z	-,339
Asymptotische Signifi- kanz (2-seitig)	,735

a Gruppenvariable: Treatment

Anhang 52: Treatmentunterschiede beim Gefallen des Gesamtbesuchs und der Stationen bei Lindenhof-Kennern. Durchgeführt wurden zuerst Kruskal-Wallis-Tests und bei signifikanten Unterschieden, d. h. $p < 0,05$, Vergleiche der Treatments mit Mann-Whitney-U-Tests

NT I:

	Lindenhof gefallen	Urwald	Sandgrube	Wachhold- erheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx
--	-----------------------	--------	-----------	----------------------	-------------------	------------------	--------------------

Chi-Quadrat	,899	,379	1,262	2,942	2,539	,039	2,189
df	2	2	2	2	2	2	2
Asymptotische Signifikanz	,638	,828	,532	,230	,281	,981	,335

a Kruskal-Wallis-Test

b Gruppenvariable: Zugang]]

Keine weiteren Tests, da keine Signifikanz

NT II:

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx
Chi-Quadrat	,294	1,222	,325	1,942	4,885	2,651	1,285
df	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,863	,543	,850	,379	,087	,266	,526

a Kruskal-Wallis-Test

b Gruppenvariable: Zugang]]

Keine weiteren Tests, da keine Signifikanz

Anhang 53: Geschlechtsdifferenzen im Gefallen des Museumsbesuchs und der Stationen der Gesamtgruppe; *AffMu*.

Ränge

	Geschlecht	N	Mittlerer Rang	Rangsumme
Lindenhof gefallen	Mädchen	169	188,26	31816,00
	Junge	178	160,46	28562,00
	Gesamt	347		
Urwald	Mädchen	174	177,94	30961,00
	Junge	182	179,04	32585,00
	Gesamt	356		
Sandgrube	Mädchen	175	181,24	31717,00
	Junge	183	177,84	32544,00
	Gesamt	358		
Wachholderheide	Mädchen	176	185,02	32563,50
	Junge	179	171,10	30626,50
	Gesamt	355		
Kulturfolger	Mädchen	173	194,88	33714,00
	Junge	179	158,74	28414,00
	Gesamt	352		
Platzbedarf	Mädchen	177	188,54	33372,00
	Junge	180	169,62	30531,00
	Gesamt	357		
Archaeopteryx	Mädchen	174	170,55	29676,50
	Junge	186	189,80	35303,50
	Gesamt	360		
Kindermuseum	Mädchen	175	188,07	32911,50
	Junge	188	176,35	33154,50
	Gesamt	363		
Lindenhof gefallen	Mädchen	174	195,53	34022,00
	Junge	173	152,35	26356,00
	Gesamt	347		
Urwald	Mädchen	171	185,81	31773,00
	Junge	182	168,73	30708,00
	Gesamt	353		

Sandgrube	Mädchen	175	184,51	32288,50
	Junge	182	173,71	31614,50
	Gesamt	357		
Wachholderheide	Mädchen	164	182,32	29900,00
	Junge	170	153,21	26045,00
	Gesamt	334		
Kulturfolger	Mädchen	171	195,32	33400,50
	Junge	175	152,17	26630,50
	Gesamt	346		
Platzbedarf	Mädchen	169	189,27	31986,00
	Junge	181	162,65	29439,00
	Gesamt	350		
Archaeopteryx	Mädchen	167	162,56	27147,00
	Junge	178	182,80	32538,00
	Gesamt	345		
Kindermuseum	Mädchen	176	191,53	33709,50
	Junge	179	164,70	29480,50
	Gesamt	355		

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archae- opteryx	Kindermu- seum
Mann- Whitney-U	12631,000	15736,000	15708,000	14516,500	12304,000	14241,000	14451,500	15388,500
Wilcoxon-W	28562,000	30961,000	32544,000	30626,500	28414,000	30531,000	29676,500	33154,500
Z	-2,932	-,108	-,334	-1,345	-3,513	-1,816	-1,840	-1,747
Asymp- totische Signifikanz (2-seitig)	,003	,914	,738	,179	,000	,069	,066	,081

a Gruppenvariable: Geschlecht

Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbedarf	Archaeop- teryx	Kindermu- seum
11305,000	14055,000	14961,500	11510,000	11230,500	12968,000	13119,000	13370,500
26356,000	30708,000	31614,500	26045,000	26630,500	29439,000	27147,000	29480,500
-4,337	-1,718	-1,041	-2,869	-4,209	-2,592	-1,975	-3,424
,000	,086	,298	,004	,000	,010	,048	,001

Ränge

	Geschlecht	N	Mittlerer Rang	Rangsumme
Lindenhof gefallen	Mädchen	58	62,01	3596,50
	Junge	60	57,08	3424,50
	Gesamt	118		
Urwald	Mädchen	56	54,91	3075,00
	Junge	59	60,93	3595,00
	Gesamt	115		
Sandgrube	Mädchen	57	57,71	3289,50
	Junge	60	60,22	3613,50
	Gesamt	117		
Wachholderheide	Mädchen	57	54,47	3105,00
	Junge	59	62,39	3681,00
	Gesamt	116		
Kulturfolger	Mädchen	57	65,31	3722,50
	Junge	59	51,92	3063,50
	Gesamt	116		
Platzbedarf	Mädchen	58	60,08	3484,50
	Junge	60	58,94	3536,50
	Gesamt	118		

Archaeopteryx	Mädchen	57	51,14	2915,00
	Junge	61	67,31	4106,00
	Gesamt	118		
Kindermuseum	Mädchen	58	60,91	3533,00
	Junge	63	61,08	3848,00
	Gesamt	121		
Lindenhof gefallen	Mädchen	57	62,93	3587,00
	Junge	60	55,27	3316,00
	Gesamt	117		
Urwald	Mädchen	57	57,88	3299,00
	Junge	60	60,07	3604,00
	Gesamt	117		
Sandgrube	Mädchen	58	59,72	3463,50
	Junge	60	59,29	3557,50
	Gesamt	118		
Wachholderheide	Mädchen	53	56,22	2979,50
	Junge	55	52,85	2906,50
	Gesamt	108		
Kulturfolger	Mädchen	57	64,84	3696,00
	Junge	57	50,16	2859,00
	Gesamt	114		
Platzbedarf	Mädchen	56	60,96	3414,00
	Junge	61	57,20	3489,00
	Gesamt	117		
Archaeopteryx	Mädchen	56	47,56	2663,50
	Junge	57	66,27	3777,50
	Gesamt	113		
Kindermuseum	Mädchen	58	63,10	3660,00
	Junge	59	54,97	3243,00
	Gesamt	117		

Anhang 54: Geschlechtsdifferenzen im Gefallen des Museumsbesuchs und der Stationen der Lindenhof-Nicht-Kenner; AffMu.

	Geschlecht	N	Mittlerer Rang	Rangsumme
Lindenhof gefallen	Mädchen	111	125,29	13907,50
	Junge	115	102,12	11743,50
	Gesamt	226		
Urwald	Mädchen	118	122,04	14400,50
	Junge	119	115,99	13802,50
	Gesamt	237		
Sandgrube	Mädchen	118	122,87	14498,50
	Junge	119	115,16	13704,50
	Gesamt	237		
Wachholderheide	Mädchen	119	128,82	15329,50
	Junge	116	106,90	12400,50
	Gesamt	235		
Kulturfolger	Mädchen	116	128,78	14938,00
	Junge	116	104,22	12090,00
	Gesamt	232		
Platzbedarf	Mädchen	119	127,12	15127,00
	Junge	116	108,65	12603,00
	Gesamt	235		
Archaeopteryx	Mädchen	117	117,21	13713,50
	Junge	121	121,71	14727,50
	Gesamt	238		
Kindermuseum	Mädchen	117	125,82	14720,50
	Junge	121	113,39	13720,50
	Gesamt	238		

Lindenhof gefallen	Mädchen	117	130,77	15300,50
	Junge	109	94,96	10350,50
	Gesamt	226		
Urwald	Mädchen	114	126,57	14429,50
	Junge	118	106,77	12598,50
	Gesamt	232		
Sandgrube	Mädchen	117	123,18	14412,50
	Junge	118	112,86	13317,50
	Gesamt	235		
Wachholderheide	Mädchen	111	124,97	13872,00
	Junge	111	98,03	10881,00
	Gesamt	222		
Kulturfolger	Mädchen	114	128,94	14699,00
	Junge	114	100,06	11407,00
	Gesamt	228		
Platzbedarf	Mädchen	113	127,12	14365,00
	Junge	116	103,19	11970,00
	Gesamt	229		
Archaeopteryx	Mädchen	111	114,03	12657,50
	Junge	117	114,94	13448,50
	Gesamt	228		
Kindermuseum	Mädchen	118	127,16	15004,50
	Junge	116	107,68	12490,50
	Gesamt	234		

Anhang 55: Geschlechtsdifferenzen im Gefallen des Museumsbesuchs und der Stationen der Lindenhof-Kenner; AffMu.

Ränge

	Geschlecht	N	Mittlerer Rang	Rangsumme
Lindenhof ge- fallen	Mädchen	58	62,01	3596,50
	Junge	60	57,08	3424,50
	Gesamt	118		
Urwald	Mädchen	56	54,91	3075,00
	Junge	59	60,93	3595,00
	Gesamt	115		
Sandgrube	Mädchen	57	57,71	3289,50
	Junge	60	60,22	3613,50
	Gesamt	117		
Wachholder- heide	Mädchen	57	54,47	3105,00
	Junge	59	62,39	3681,00
	Gesamt	116		
Kulturfolger	Mädchen	57	65,31	3722,50
	Junge	59	51,92	3063,50
	Gesamt	116		
Platzbedarf	Mädchen	58	60,08	3484,50
	Junge	60	58,94	3536,50
	Gesamt	118		
Archaeopteryx	Mädchen	57	51,14	2915,00
	Junge	61	67,31	4106,00
	Gesamt	118		
Kindermuseum	Mädchen	58	60,91	3533,00
	Junge	63	61,08	3848,00
	Gesamt	121		
Lindenhof ge- fallen	Mädchen	57	62,93	3587,00
	Junge	60	55,27	3316,00

	Gesamt	117		
Urwald	Mädchen	57	57,88	3299,00
	Junge	60	60,07	3604,00
	Gesamt	117		
Sandgrube	Mädchen	58	59,72	3463,50
	Junge	60	59,29	3557,50
	Gesamt	118		
Wachholder- heide	Mädchen	53	56,22	2979,50
	Junge	55	52,85	2906,50
	Gesamt	108		
Kulturfolger	Mädchen	57	64,84	3696,00
	Junge	57	50,16	2859,00
	Gesamt	114		
Platzbedarf	Mädchen	56	60,96	3414,00
	Junge	61	57,20	3489,00
	Gesamt	117		
Archaeopteryx	Mädchen	56	47,56	2663,50
	Junge	57	66,27	3777,50
	Gesamt	113		
Kindermuseum	Mädchen	58	63,10	3660,00
	Junge	59	54,97	3243,00
	Gesamt	117		

Statistik für Test

	Linden- hof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Mann- Whitney-U	1594,500	1479,000	1636,500	1452,000	1293,500	1706,500	1262,000	1822,000
Wilcoxon- W	3424,500	3075,000	3289,500	3105,000	3063,500	3536,500	2915,000	3533,000
Z	-,869	-1,033	-,431	-1,338	-2,271	-,190	-2,685	-,043
Asymp- totische Signifikanz (2-seitig)	,385	,301	,667	,181	,023	,850	,007	,966

a Gruppenvariable: Geschlecht

	Linden- hof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Mann- Whitney-U	1486,000	1646,000	1727,500	1366,500	1206,000	1598,000	1067,500	1473,000
Wilcoxon- W	3316,000	3299,000	3557,500	2906,500	2859,000	3489,000	2663,500	3243,000
Z	-1,349	-,385	-,071	-,584	-2,490	-,628	-3,204	-1,698
Asymp- totische Signifikanz (2-seitig)	,177	,701	,943	,559	,013	,530	,001	,089

Anhang 56: Affektive Treatmentunterschiede bei Jungen in Gesamtgruppe und den beiden Teilgruppen (Nicht-Kenner, Kenner); AffMu.

Ränge

	Treatment	N	Mittlerer Rang
Linden- hof ge- fallen	S	54	86,21
	SF	64	92,59
	F	60	89,17
	Gesamt	178	

Urwald	S	59	84,75
	SF	64	92,50
	F	59	97,16
	Gesamt	182	
Sandgrube	S	59	88,42
	SF	65	94,76
	F	59	92,53
	Gesamt	183	
Wachholder- heide	S	59	78,44
	SF	62	87,15
	F	58	104,80
	Gesamt	179	
Kulturfolger	S	58	90,85
	SF	60	81,23
	F	61	97,81
	Gesamt	179	
Platzbedarf	S	58	81,26
	SF	64	87,41
	F	58	103,15
	Gesamt	180	
Archaeopteryx	S	60	93,54
	SF	66	82,77
	F	60	105,27
	Gesamt	186	
Kindermuseum	S	60	95,47
	SF	67	94,10
	F	61	93,98
	Gesamt	188	
Lindenhof ge- fallen	S	58	83,23
	SF	64	91,43
	F	51	85,73
	Gesamt	173	
Urwald	S	59	82,83
	SF	67	94,37
	F	56	97,20
	Gesamt	182	
Sandgrube	S	59	89,36
	SF	67	87,70
	F	56	98,30
	Gesamt	182	
Wachholder- heide	S	55	88,01
	SF	60	83,41
	F	55	85,27
	Gesamt	170	
Kulturfolger	S	57	91,92
	SF	64	88,28
	F	54	83,53
	Gesamt	175	
Platzbedarf	S	59	87,08
	SF	65	80,75
	F	57	106,74
	Gesamt	181	
Archaeopteryx	S	56	85,75
	SF	65	93,55
	F	57	88,57
	Gesamt	178	

Kindermuseum	S	59	92,38
	SF	65	89,42
	F	55	88,14
	Gesamt	179	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhold- erheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,549	1,877	,514	8,645	3,425	5,972	6,169	,064
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,760	,391	,774	,013	,180	,051	,046	,969

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,982	2,839	1,514	,271	,832	8,788	,798	,320
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,612	,242	,469	,873	,660	,012	,671	,852

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Wachholderheide	S	59	57,61	3399,00
	SF	62	64,23	3982,00
	Gesamt	121		
Platzbedarf	S	58	59,30	3439,50
	SF	64	63,49	4063,50
	Gesamt	122		
Archaeopteryx	S	60	67,13	4027,50
	SF	66	60,20	3973,50
	Gesamt	126		
Platzbedarf	S	59	64,88	3828,00
	SF	65	60,34	3922,00
	Gesamt	124		

Statistik für Test

	Wachhol- derheide	Platzbedarf	Archaeop- teryx	Platzbedarf
Mann- Whitney-U	1629,000	1728,500	1762,500	1777,000
Wilcoxon- W	3399,000	3439,500	3973,500	3922,000
Z	-1,096	-,684	-1,115	-,741
Asymp- totische Signifikanz (2-seitig)	,273	,494	,265	,459

a Gruppenvariable: Treatment

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Wachholder- heide	S	59	50,83	2999,00

	F	58	67,31	3904,00
	Gesamt	117		
Platzbedarf	S	58	51,46	2984,50
	F	58	65,54	3801,50
	Gesamt	116		
Archaeop- teryx	S	60	56,92	3415,00
	F	60	64,08	3845,00
	Gesamt	120		
Platzbedarf	S	59	52,20	3080,00
	F	57	65,02	3706,00
	Gesamt	116		

Statistik für Test

	Wachhold erheide	Platzbe- darf	Archaeop- teryx	Platzbe- darf
Mann- Whitney-U	1229,000	1273,500	1585,000	1310,000
Wilcoxon- W	2999,000	2984,500	3415,000	3080,000
Z	-2,738	-2,357	-1,217	-2,159
Asymp- totische Signifikanz (2-seitig)	,006	,018	,224	,031

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rang- summe
Wachhold erheide	SF	62	54,43	3374,50
	F	58	66,99	3885,50
	Gesamt	120		
Platzbe- darf	SF	64	56,42	3611,00
	F	58	67,10	3892,00
	Gesamt	122		
Archaeop- teryx	SF	66	56,06	3700,00
	F	60	71,68	4301,00
	Gesamt	126		
Platzbe- darf	SF	65	53,42	3472,00
	F	57	70,72	4031,00
	Gesamt	122		

Statistik für Test

	Wachhold erheide	Platzbedarf	Archaeop- teryx	Platzbedarf
Mann- Whitney-U	1421,500	1531,000	1489,000	1327,000
Wilcoxon- W	3374,500	3611,000	3700,000	3472,000
Z	-2,083	-1,746	-2,540	-2,829
Asymp- totische Signifikanz (2-seitig)	,037	,081	,011	,005

a Gruppenvariable: Treatment

Lindenhof-Nicht-Kenner nur Jungen

Ränge

	Treatment	N	Mittlerer Rang
Lindenhof gefallen	S	34	54,79
	SF	43	61,33
	F	38	57,11
	Gesamt	115	
Urwald	S	35	51,86
	SF	46	62,07
	F	38	65,00
	Gesamt	119	
Sandgrube	S	35	60,90
	SF	46	59,91
	F	38	59,28
	Gesamt	119	
Wachhold erheide	S	36	50,57
	SF	44	57,63
	F	36	67,50
	Gesamt	116	
Kulturfolger	S	35	59,84
	SF	42	52,49
	F	39	63,77
	Gesamt	116	
Platzbedarf	S	35	51,83
	SF	44	56,72
	F	37	66,93
	Gesamt	116	
Archaeopteryx	S	36	65,10
	SF	46	54,43
	F	39	64,96
	Gesamt	121	
Kindermuseum	S	36	61,28
	SF	46	61,02
	F	39	60,72
	Gesamt	121	
Lindenhof gefallen	S	35	50,77
	SF	43	57,62
	F	31	56,15
	Gesamt	109	
Urwald	S	36	52,43
	SF	46	62,62
	F	36	62,58
	Gesamt	118	
Sandgrube	S	36	56,33
	SF	46	57,80
	F	36	64,83
	Gesamt	118	
Wachhold erheide	S	34	57,38
	SF	42	55,60

	F	35	55,14
	Gesamt	111	
Kulturfolger	S	35	58,04
	SF	44	57,80
	F	35	56,59
	Gesamt	114	
Platzbedarf	S	36	59,06
	SF	44	50,23
	F	36	68,06
	Gesamt	116	
Archaeopteryx	S	35	54,47
	SF	45	61,94
	F	37	59,70
	Gesamt	117	
Kindermuseum	S	36	59,46
	SF	44	58,57
	F	36	57,46
	Gesamt	116	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,937	3,273	,048	5,094	2,560	4,208	2,925	,010
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,626	,195	,977	,078	,278	,122	,232	,995

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	1,089	2,526	1,411	,101	,043	6,182	1,072	,102
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,580	,283	,494	,951	,979	,045	,585	,950

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Platzbedarf	S	36	43,89	1580,00
	SF	44	37,73	1660,00
	Gesamt	80		

Statistik für Test

	Platzbedarf
Mann- Whitney-U	670,000
Wilcoxon- W	1660,000
Z	-1,258
Asymp-	,209

totische Signifikanz (2-seitig)	
---------------------------------------	--

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rang- summe
Platzbe- darf	S	36	33,67	1212,00
	F	36	39,33	1416,00
	Gesamt	72		

Statistik für Test

	Platzbedarf
Mann- Whitney-U	546,000
Wilcoxon- W	1212,000
Z	-1,202
Asymp- totische Signifikanz (2-seitig)	,229

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rang- summe
Platzbe- darf	SF	44	35,00	1540,00
	F	36	47,22	1700,00
	Gesamt	80		

Statistik für Test

	Platzbe- darf
Mann- Whitney-U	550,000
Wilcoxon- W	1540,000
Z	-2,459
Asymp- totische Signifikanz (2-seitig)	,014

a Gruppenvariable: Treatment

Lindenhof-Kenner nur Jungen

Ränge

	Treatment	N	Mittlerer Rang
Lindenhof gefallen	S	18	29,00
	SF	20	30,90
	F	22	31,36
	Gesamt	60	
Urwald	S	21	29,67

	SF	17	29,00
	F	21	31,14
	Gesamt	59	
Sandgrube	S	21	26,05
	SF	18	33,44
	F	21	32,43
	Gesamt	60	
Wachhol- derheide	S	20	24,67
	SF	17	29,12
	F	22	35,52
	Gesamt	59	
Kulturfol- ger	S	20	26,80
	SF	17	30,00
	F	22	32,91
	Gesamt	59	
Platzbe- darf	S	20	25,98
	SF	19	31,21
	F	21	34,17
	Gesamt	60	
Archaeop- teryx	S	21	26,98
	SF	19	27,63
	F	21	38,07
	Gesamt	61	
Kindermu- seum	S	21	32,07
	SF	20	31,85
	F	22	32,07
	Gesamt	63	
Lindenhof gefallen	S	20	30,60
	SF	20	31,80
	F	20	29,10
	Gesamt	60	
Urwald	S	20	27,83
	SF	20	30,10
	F	20	33,58
	Gesamt	60	
Sandgrube	S	20	30,50
	SF	20	29,15
	F	20	31,85
	Gesamt	60	
Wachhold erheide	S	18	28,44
	SF	17	26,79
	F	20	28,63
	Gesamt	55	
Kulturfol- ger	S	19	30,79
	SF	19	30,16
	F	19	26,05
	Gesamt	57	
Platzbe- darf	S	20	26,33
	SF	20	29,13
	F	21	37,24

	Gesamt	61	
Archaeopteryx	S	18	28,22
	SF	19	31,82
	F	20	27,02
	Gesamt	57	
Kindermuseum	S	20	30,75
	SF	20	29,45
	F	19	29,79
	Gesamt	59	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,237	,176	2,350	4,657	1,553	2,533	5,769	,005
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,888	,916	,309	,097	,460	,282	,056	,997

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,275	1,285	,268	,154	,997	4,598	1,024	,090
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,871	,526	,875	,926	,608	,100	,599	,956

Ränge

	Treatment	N	Mittlerer Rang	Rang- summe
Archaeop- teryx	S	21	20,17	423,50
	SF	19	20,87	396,50
	Gesamt	40		

Statistik für Test

	Archaeop- teryx
Mann- Whitney-U	192,500
Wilcoxon- W	423,500
Z	-,199
Asymp- totische Signifikanz (2-seitig)	,842
Exakte Signifikanz [2*(1-seitig Sig.)]	,851

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Archaeopteryx	S	21	17,81	374,00
	F	21	25,19	529,00
	Gesamt	42		

Statistik für Test

	Archaeopteryx
Mann-Whitney-U	143,000
Wilcoxon-W	374,000
Z	-2,097
Asymptotische Signifikanz (2-seitig)	,036

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Archaeopteryx	SF	19	16,76	318,50
	F	21	23,88	501,50
	Gesamt	40		

Statistik für Test

	Archaeopteryx
Mann-Whitney-U	128,500
Wilcoxon-W	318,500
Z	-2,076
Asymptotische Signifikanz (2-seitig)	,038
Exakte Signifikanz [2*(1-seitig Sig.)]	,054

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Treatment

Anhang 57: Affektive Treatmentunterschiede bei Mädchen in Gesamtgruppe und den beiden Teilgruppen (Nicht-Kenner, Kenner); AffMu.

Ränge

	Treatment	N	Mittlerer Rang
Lindenhof gefallen	S	54	86,21
	SF	64	92,59

	F	60	89,17
	Gesamt	178	
Urwald	S	59	84,75
	SF	64	92,50
	F	59	97,16
	Gesamt	182	
Sandgrube	S	59	88,42
	SF	65	94,76
	F	59	92,53
	Gesamt	183	
Wachhold erheide	S	59	78,44
	SF	62	87,15
	F	58	104,80
	Gesamt	179	
Kulturfol- ger	S	58	90,85
	SF	60	81,23
	F	61	97,81
	Gesamt	179	
Platzbe- darf	S	58	81,26
	SF	64	87,41
	F	58	103,15
	Gesamt	180	
Archaeop- teryx	S	60	93,54
	SF	66	82,77
	F	60	105,27
	Gesamt	186	
Kindermu- seum	S	60	95,47
	SF	67	94,10
	F	61	93,98
	Gesamt	188	
Lindenhof gefallen	S	58	83,23
	SF	64	91,43
	F	51	85,73
	Gesamt	173	
Urwald	S	59	82,83
	SF	67	94,37
	F	56	97,20
	Gesamt	182	
Sandgrube	S	59	89,36
	SF	67	87,70
	F	56	98,30
	Gesamt	182	
Wachhold erheide	S	55	88,01
	SF	60	83,41
	F	55	85,27
	Gesamt	170	
Kulturfol- ger	S	57	91,92
	SF	64	88,28
	F	54	83,53
	Gesamt	175	
Platzbe-	S	59	87,08

darf			
	SF	65	80,75
	F	57	106,74
	Gesamt	181	
Archaeop- teryx	S	56	85,75
	SF	65	93,55
	F	57	88,57
	Gesamt	178	
Kindermu- seum	S	59	92,38
	SF	65	89,42
	F	55	88,14
	Gesamt	179	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx	Kindermu- seum
Chi- Quadrat	,549	1,877	,514	8,645	3,425	5,972	6,169	,064
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,760	,391	,774	,013	,180	,051	,046	,969

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archae- opteryx	Kindermu- seum
Chi- Quadrat	,982	2,839	1,514	,271	,832	8,788	,798	,320
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,612	,242	,469	,873	,660	,012	,671	,852

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Wachholder heide	S	59	57,61	3399,00
	SF	62	64,23	3982,00
	Gesamt	121		
Platzbedarf	S	58	59,30	3439,50
	SF	64	63,49	4063,50
	Gesamt	122		
Archaeop- teryx	S	60	67,13	4027,50
	SF	66	60,20	3973,50
	Gesamt	126		
Platzbedarf	S	59	64,88	3828,00
	SF	65	60,34	3922,00
	Gesamt	124		

Statistik für Test

	Wachhol- derheide	Platzbe- darf	Archaeop- teryx	Platzbe- darf
Mann- Whitney-U	1629,000	1728,500	1762,500	1777,000
Wilcoxon-	3399,000	3439,500	3973,500	3922,000

W				
Z	-1,096	-,684	-1,115	-,741
Asymp- totische Signifikanz (2-seitig)	,273	,494	,265	,459

a Gruppenvariable: Treatment

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Wachholder- heide	S	59	50,83	2999,00
	F	58	67,31	3904,00
	Gesamt	117		
Platzbedarf	S	58	51,46	2984,50
	F	58	65,54	3801,50
	Gesamt	116		
Archaeop- teryx	S	60	56,92	3415,00
	F	60	64,08	3845,00
	Gesamt	120		
Platzbedarf	S	59	52,20	3080,00
	F	57	65,02	3706,00
	Gesamt	116		

Statistik für Test

	Wachhold- erheide	Platzbedarf	Archaeop- teryx	Platzbedarf
Mann- Whitney-U	1229,000	1273,500	1585,000	1310,000
Wilcoxon- W	2999,000	2984,500	3415,000	3080,000
Z	-2,738	-2,357	-1,217	-2,159
Asymp- totische Signifikanz (2-seitig)	,006	,018	,224	,031

a Gruppenvariable: Treatment

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Wachholder- heide	SF	62	54,43	3374,50
	F	58	66,99	3885,50
	Gesamt	120		
Platzbedarf	SF	64	56,42	3611,00
	F	58	67,10	3892,00
	Gesamt	122		
Archaeop- teryx	SF	66	56,06	3700,00
	F	60	71,68	4301,00
	Gesamt	126		
Platzbedarf	SF	65	53,42	3472,00
	F	57	70,72	4031,00
	Gesamt	122		

Statistik für Test

	Wachhol- derheide	Platzbe- darf	Archaeop- teryx	Platzbe- darf
Mann- Whitney-U	1421,500	1531,000	1489,000	1327,000
Wilcoxon- W	3374,500	3611,000	3700,000	3472,000
Z	-2,083	-1,746	-2,540	-2,829
Asymp- totische Signifikanz (2-seitig)	,037	,081	,011	,005

a Gruppenvariable: Treatment

Lindenhof-Nicht-Kenner nur Mädchen
Ränge

	Treat- ment	N	Mittlerer Rang
Lindenhof gefallen	S	42	51,35
	SF	33	58,47
	F	36	59,17
	Gesamt	111	
Urwald	S	44	53,03
	SF	35	63,57
	F	39	63,14
	Gesamt	118	
Sandgrube	S	44	55,48
	SF	34	64,63
	F	40	59,56
	Gesamt	118	
Wachholder- heide	S	44	59,89
	SF	35	65,26
	F	40	55,53
	Gesamt	119	
Kulturfolger	S	44	59,72
	SF	33	57,02
	F	39	58,38
	Gesamt	116	
Platzbedarf	S	44	59,17
	SF	35	54,69
	F	40	65,56
	Gesamt	119	
Archaeop- teryx	S	44	56,28
	SF	34	69,85
	F	39	52,60
	Gesamt	117	
Kindermu- seum	S	43	59,05
	SF	34	57,79
	F	40	59,97
	Gesamt	117	
Lindenhof gefallen	S	44	57,23
	SF	33	60,35
	F	40	59,84
	Gesamt	117	

Urwald	S	44	47,68
	SF	33	65,88
	F	37	61,70
	Gesamt	114	
Sandgrube	S	44	52,06
	SF	33	69,47
	F	40	58,00
	Gesamt	117	
Wachholderheide	S	42	51,44
	SF	32	69,45
	F	37	49,54
	Gesamt	111	
Kulturfolger	S	44	56,20
	SF	33	54,17
	F	37	62,01
	Gesamt	114	
Platzbedarf	S	43	52,97
	SF	33	55,39
	F	37	63,12
	Gesamt	113	
Archaeopteryx	S	42	56,33
	SF	32	63,94
	F	37	48,76
	Gesamt	111	
Kindermuseum	S	44	62,38
	SF	34	56,38
	F	40	58,99
	Gesamt	118	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	2,198	2,887	1,580	1,690	,141	2,102	5,704	,297
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,333	,236	,454	,430	,932	,350	,058	,862

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfol- ger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,253	8,117	5,709	8,772	1,229	2,319	4,207	1,911
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,881	,017	,058	,012	,541	,314	,122	,385

Ränge

Statistik für Test

	Archaeop- teryx	Urwald	Sandgrube	Wachhol- derheide
Mann- Whitney-U	579,000	496,000	520,000	450,500

Wilcoxon-W	1569,000	1486,000	1510,000	1353,500
Z	-1,807	-2,576	-2,242	-2,552
Asymptotische Signifikanz (2-seitig)	,071	,010	,025	,011

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Archaeopteryx	S	44	43,13	1897,50
	F	39	40,73	1588,50
	Gesamt	83		
Urwald	S	44	36,41	1602,00
	F	37	46,46	1719,00
	Gesamt	81		
Sandgrube	S	44	40,24	1770,50
	F	40	44,99	1799,50
	Gesamt	84		
Wachholderheide	S	42	40,71	1710,00
	F	37	39,19	1450,00
	Gesamt	79		

Statistik für Test

	Archaeopteryx	Urwald	Sandgrube	Wachholderheide
Mann-Whitney-U	808,500	612,000	780,500	747,000
Wilcoxon-W	1588,500	1602,000	1770,500	1450,000
Z	-,472	-2,093	-,952	-,311
Asymptotische Signifikanz (2-seitig)	,637	,036	,341	,756

a Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Archaeopteryx	SF	34	42,88	1458,00
	F	39	31,87	1243,00
	Gesamt	73		
Urwald	SF	33	36,91	1218,00
	F	37	34,24	1267,00
	Gesamt	70		
Sandgrube	SF	33	41,23	1360,50
	F	40	33,51	1340,50
	Gesamt	73		
Wachholderheide	SF	32	41,53	1329,00
	F	37	29,35	1086,00
	Gesamt	69		

Statistik für Test

	Archaeop- teryx	Urwald	Sandgrube	Wachhol- derheide
Mann-Whitney-U	463,000	564,000	520,500	383,000
Wilcoxon-W	1243,000	1267,000	1340,500	1086,000
Z	-2,342	-,643	-1,668	-2,644
Asymp- totische Signifikanz (2-seitig)	,019	,520	,095	,008

a Gruppenvariable: Treatment

Lindenhof-Kenner nur Mädchen

Ränge

	Treat- ment	N	Mittlerer Rang
Lindenhof gefallen	S	19	33,95
	SF	18	29,33
	F	21	25,62
	Gesamt	58	
Urwald	S	18	31,00
	SF	18	26,86
	F	20	27,73
	Gesamt	56	
Sandgrube	S	19	31,13
	SF	18	30,78
	F	20	25,38
	Gesamt	57	
Wachholder heide	S	19	30,71
	SF	18	26,14
	F	20	29,95
	Gesamt	57	
Kulturfolger	S	19	37,84
	SF	18	24,11
	F	20	25,00
	Gesamt	57	
Platzbedarf	S	19	33,76
	SF	18	28,47
	F	21	26,52
	Gesamt	58	
Archaeop- teryx	S	19	35,58
	SF	18	26,36
	F	20	25,13
	Gesamt	57	
Kindermu- seum	S	19	33,50
	SF	18	30,28
	F	21	25,21
	Gesamt	58	
Lindenhof gefallen	S	19	31,00
	SF	17	27,47
	F	21	28,43
	Gesamt	57	
Urwald	S	19	33,95

	SF	18	24,47
	F	20	28,38
	Gesamt	57	
Sandgrube	S	19	32,00
	SF	18	29,72
	F	21	27,05
	Gesamt	58	
Wachhol- derheide	S	18	31,67
	SF	18	23,50
	F	17	25,76
	Gesamt	53	
Kulturfolger	S	19	36,42
	SF	18	22,19
	F	20	28,08
	Gesamt	57	
Platzbedarf	S	19	31,00
	SF	18	25,44
	F	19	28,89
	Gesamt	56	
Archaeop- teryx	S	18	32,44
	SF	18	27,42
	F	20	25,92
	Gesamt	56	
Kindermu- seum	S	19	33,55
	SF	18	30,14
	F	21	25,29
	Gesamt	58	

Statistik für Test

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	3,094	,773	1,869	,927	9,260	2,160	5,033	6,847
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,213	,679	,393	,629	,010	,340	,081	,033

a Kruskal-Wallis-Test

b Gruppenvariable: Treatment

	Lindenhof gefallen	Urwald	Sandgrube	Wachhol- derheide	Kulturfolger	Platzbe- darf	Archaeop- teryx	Kindermu- seum
Chi- Quadrat	,597	3,903	,957	2,922	7,790	1,192	1,779	5,217
df	2	2	2	2	2	2	2	2
Asymp- totische Signifikanz	,742	,142	,620	,232	,020	,551	,411	,074

Ränge

	Treatment	N	Mittlerer Rang	Rang- summe
Kulturfol- ger	S	19	23,34	443,50
	SF	18	14,42	259,50
	Gesamt	37		
Kindermu-	S	19	20,00	380,00

seum				
	SF	18	17,94	323,00
	Gesamt	37		
Kulturfolger	S	19	23,55	447,50
	SF	18	14,19	255,50
	Gesamt	37		

Statistik für Test

	Kulturfolger	Kindermuseum	Kulturfolger
Mann-Whitney-U	88,500	152,000	84,500
Wilcoxon-W	259,500	323,000	255,500
Z	-2,734	-1,473	-2,815
Asymptotische Signifikanz (2-seitig)	,006	,141	,005
Exakte Signifikanz [2*(1-seitig Sig.)]	,011	,578	,007

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Treatment

Ränge

	Treatment	N	Mittlerer Rang	Rangsumme
Kulturfolger	S	19	24,50	465,50
	F	20	15,73	314,50
	Gesamt	39		
Kindermuseum	S	19	23,50	446,50
	F	21	17,79	373,50
	Gesamt	40		
Kulturfolger	S	19	22,87	434,50
	F	20	17,27	345,50
	Gesamt	39		

Statistik für Test

	Kulturfolger	Kinder-museum	Kulturfolger
Mann-Whitney-U	104,500	142,500	135,500
Wilcoxon-W	314,500	373,500	345,500
Z	-2,628	-2,487	-1,654
Asymptotische Signifikanz (2-seitig)	,009	,013	,098
Exakte Signifikanz [2*(1-seitig Sig.)]	,015	,124	,127

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Treatment

Ränge

	Treat- ment	N	Mittlerer Rang	Rang- summe
Kulturfolger	SF	18	19,19	345,50
	F	20	19,77	395,50
	Gesamt	38		
Kindermu- seum	SF	18	21,83	393,00
	F	21	18,43	387,00
	Gesamt	39		
Kulturfolger	SF	18	17,50	315,00
	F	20	21,30	426,00
	Gesamt	38		

Statistik für Test

	Kulturfol- ger	Kindermu- seum	Kulturfolger
Mann-Whitney-U	174,500	156,000	144,000
Wilcoxon-W	345,500	387,000	315,000
Z	-,170	-1,320	-1,108
Asymp- totische Signifikanz (2-seitig)	,865	,187	,268
Exakte Signifikanz [2*(1-seitig Sig.)]	,874	,364	,303

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Treatment

Anhang 58: Treatmentunterschiede

Gesamtgruppe

Sprache Hauptinhalt:bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,654$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,100$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,752$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,442$ **Zeichnung Hauptinhalt:**bei Mädchen / Nachtest I: Exakter Fisher-Test: p nicht zu berechnenbei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,915$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,414$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,116$ **Zeichnung ja / nein:**bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,326$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,743$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,366$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,016$

Kreuztabelle

			Zugang			Gesamt
			konstruktiv- istisch	mäßig instruk- tivistisch	Instruktivistisch	
Zeichnung	nein	Anzahl	39	28	38	105
		Erwartete An- zahl	33,5	37,4	34,1	105,0
	unordentlich	Anzahl	21	39	23	83
		Erwartete An- zahl	26,5	29,6	26,9	83,0
Gesamt		Anzahl	60	67	61	188
		Erwartete An- zahl	60,0	67,0	61,0	188,0

Lindenhof-Nicht-Kenner

Sprache Hauptinhalt:bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,940$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,025$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,096$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,217$

Kreuztabelle

			Zugang			Gesamt
			konstruktiv- istisch	mäßig in- struktivistisch	Instruktivistisch	
Bemerkung	keine Antwort	Anzahl	14	12	4	30
		Erwartete Anzahl	11,1	8,8	10,1	30,0
	pos. allg. Wertung	Anzahl	27	22	35	84
		Erwartete Anzahl	31,1	24,7	28,2	84,0
	Pos+neg. Wertung	Anzahl	2	0	0	2
		Erwartete Anzahl	,7	,6	,7	2,0
	krit. An- merkung/Vors- chlag	Anzahl	1	1	1	3
		Erwartete Anzahl	1,1	,9	1,0	3,0
Gesamt		Anzahl	44	35	40	119
		Erwartete Anzahl	44,0	35,0	40,0	119,0

Zeichnung Hauptinhalt:bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,034$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,833$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,435$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,577$

Kreuztabelle

			Zugang			Gesamt
			konstruktiv- istisch	mäßig in- struktivistisch	Instruktivistisch	
Hauptinhalt	nichts	Anzahl	7	2	5	14
		Erwartete	5,2	4,1	4,7	14,0

		Anzahl				
	Kindermuseum	Anzahl	23	12	14	49
		Erwartete Anzahl	18,1	14,4	16,5	49,0
	Naturkundemuseum	Anzahl	5	8	2	15
		Erwartete Anzahl	5,5	4,4	5,0	15,0
	beide	Anzahl	5	3	7	15
		Erwartete Anzahl	5,5	4,4	5,0	15,0
	nicht Lindenhof	Anzahl	1	7	7	15
		Erwartete Anzahl	5,5	4,4	5,0	15,0
	Zwergmaus	Anzahl	3	1	5	9
		Erwartete Anzahl	3,3	2,6	3,0	9,0
	nicht erkennbar	Anzahl	0	1	0	1
		Erwartete Anzahl	,4	,3	,3	1,0
	Betreuerin	Anzahl	0	1	0	1
		Erwartete Anzahl	,4	,3	,3	1,0
Gesamt		Anzahl	44	35	40	119
		Erwartete Anzahl	44,0	35,0	40,0	119,0

Zeichnung ja / nein:bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,375$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,285$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,813$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,078$

Lindenhof-Kenner

Sprache Hauptinhalt:bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,710$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,917$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,054$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,559$ **Zeichnung Hauptinhalt:**bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,171$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,874$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,235$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,471$ **Zeichnung ja / nein:**bei Mädchen / Nachtest I: Exakter Fisher-Test: $p=0,076$ bei Mädchen / Nachtest II: Exakter Fisher-Test: $p=0,760$ bei Jungen / Nachtest I: Exakter Fisher-Test: $p=0,547$ bei Jungen / Nachtest II: Exakter Fisher-Test: $p=0,290$

Anhang 59: Tabellenanhang:*KogMit18* alle Schüler

(1.1.1.1.)

	Anzahl	Mittelwert	Standard- abweichung	Median	25. Perzentil	75. Perzentil
Richtig VT	366	5,65	3,03	5,00	3,00	8,00
Richtig NT I	366	9,01	3,00	9,00	7,00	11,00
Richtig NT II	366	8,07	3,40	8,00	6,00	10,00

(1.1.1.2.) (1.1.1.3.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Richtig VT	124	5,85	3,25	5,00	3,00	8,75
		Richtig NT I	124	8,07	3,20	8,00	6,00	10,00
		Richtig NT II	124	7,45	3,47	7,50	5,00	10,00
	SF	Richtig VT	120	5,63	2,96	5,00	3,00	8,00
		Richtig NT I	120	9,90	2,66	10,00	9,00	12,00
		Richtig NT II	120	8,86	3,38	9,00	7,00	11,00
	F	Richtig VT	122	5,46	2,90	5,00	3,00	8,00
		Richtig NT I	122	9,08	2,84	9,00	7,00	11,00
		Richtig NT II	122	7,93	3,20	8,00	6,00	10,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Perzentil
Richtige F-Fragen VT	366	2,45	1,38	2,00	1,00	3,00
Richtige F-Fragen NT I	366	3,54	1,39	4,00	2,00	5,00
Richtige F-Fragen NT II	366	2,96	1,37	3,00	2,00	4,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Richtige SF-Fragen VT	366	1,54	1,29	1,00	1,00	2,00
Richtige SF-Fragen NT I	366	2,63	1,47	3,00	2,00	4,00
Richtige SF-Fragen NT II	366	2,43	1,46	2,00	1,00	3,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Richtige S-Fragen VT	366	1,66	1,20	1,00	1,00	2,00
Richtige S-Fragen NT I	366	2,84	1,26	3,00	2,00	4,00
Richtige S-Fragen NT II	366	2,68	1,40	3,00	2,00	4,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige S-Fragen VT	124	1,74	1,23	2,00	1,00	2,00
		Richtige S-Fragen NT I	124	2,81	1,32	3,00	2,00	4,00
		Richtige S-Fragen NT II	124	2,76	1,53	3,00	2,00	4,00
	SF	Richtige S-Fragen VT	120	1,66	1,23	1,00	1,00	3,00
		Richtige S-Fragen NT I	120	2,87	1,23	3,00	2,00	4,00
		Richtige S-Fragen NT II	120	2,78	1,36	3,00	2,00	4,00
	F	Richtige S-Fragen VT	122	1,59	1,16	1,00	1,00	2,00
		Richtige S-Fragen NT I	122	2,84	1,25	3,00	2,00	4,00
		Richtige S-Fragen NT II	122	2,51	1,30	2,00	2,00	3,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige SF-Fragen VT	124	1,61	1,36	1,00	1,00	3,00
		Richtige SF-Fragen NT I	124	2,17	1,38	2,00	1,00	3,00
		Richtige SF-Fragen NT II	124	2,02	1,41	2,00	1,00	3,00
	SF	Richtige SF-Fragen VT	120	1,53	1,28	1,00	1,00	2,00
		Richtige SF-Fragen NT I	120	3,56	1,19	4,00	3,00	4,00
		Richtige SF-Fragen NT II	120	2,95	1,40	3,00	2,00	4,00
	F	Richtige SF-Fragen VT	122	1,48	1,24	1,00	,00	2,00
		Richtige SF-Fragen NT I	122	2,17	1,38	2,00	1,00	3,00
		Richtige SF-Fragen NT II	122	2,33	1,42	2,00	1,00	3,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige F-Fragen VT	124	2,49	1,39	2,00	1,00	3,00
		Richtige F-Fragen NT I	124	3,09	1,43	3,00	2,00	4,00
		Richtige F-Fragen NT II	124	2,68	1,27	3,00	2,00	4,00
	SF	Richtige F-Fragen VT	120	2,45	1,31	2,00	1,25	3,00

		Richtige F-Fragen NT II	120	3,13	1,42	3,00	2,00	4,00
	F	Richtige F-Fragen VT	122	2,39	1,43	2,00	1,00	4,00
		Richtige F-Fragen NT I	122	4,07	1,31	4,00	3,00	5,00
		Richtige F-Fragen NT II	122	3,09	1,40	3,00	2,00	4,00

KogO alle Schüler
(1.1.2.1.)

	Anzahl	Mittelwert	Standard- abweichung	Median	25. Per- zentil	75. Per- zentil
SummenscoreVT	366	5,24	3,74	5,00	3,00	7,00
SummenscoreNT I	366	10,65	5,74	10,00	6,00	14,00
SummenscoreNT II	366	9,34	6,07	9,00	5,00	13,00

(1.1.2.2.) (1.1.2.3.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Summenscore VT	124	5,38	3,99	5,00	3,00	7,00
		Summenscore NT I	124	12,27	5,90	12,00	8,25	15,75
		Summenscore NT II	124	10,04	6,16	9,00	6,00	15,00
	SF	Summenscore VT	120	5,15	3,38	5,00	3,00	7,00
		Summenscore NT I	120	9,83	5,20	10,00	6,00	13,00
		Summenscore NT II	120	8,80	5,81	9,00	4,00	12,00
	F	Summenscore VT	122	5,20	3,83	4,50	3,00	6,25
		Summenscore NT I	122	9,81	5,78	9,50	6,00	14,00
		Summenscore NT II	122	9,17	6,20	9,00	4,00	13,00

KogMit18
(1.2.1.1.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Lindenhof bekannt	ja	RichtigVT	121	6,17	3,25	6,00	4,00	9,00
		Richtig NT I	121	9,54	3,03	10,00	7,00	12,00
		Richtig NT II	121	9,00	3,44	9,00	7,00	11,00
	nein	Richtig VT	241	5,36	2,89	5,00	3,00	7,00
		Richtig NT I	241	8,72	2,94	9,00	7,00	11,00
		Richtig NT II	241	7,59	3,29	8,00	6,00	10,00

KogMit18 Lindenhof-Nicht-Kenner
(1.2.1.1.1.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Richtig VT	81	5,40	2,88	5,00	3,00	7,00
		Richtig NT I	81	7,49	2,89	8,00	6,00	9,00
		Richtig NT II	81	6,93	3,20	7,00	5,00	9,00
	SF	Richtig VT	81	5,37	2,86	5,00	3,00	7,00
		Richtig NT I	81	9,80	2,69	10,00	9,00	12,00
		Richtig NT II	81	8,49	3,60	9,00	6,00	11,50
	F	Richtig VT	79	5,32	2,97	5,00	3,00	8,00
		Richtig NT I	79	8,86	2,82	9,00	7,00	11,00
		Richtig NT II	79	7,34	2,85	8,00	5,00	9,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Richtige S-Fragen VT	241	1,60	1,11	1,00	1,00	2,00
Richtige S-Fragen NT I	241	2,68	1,18	3,00	2,00	3,50
Richtige S-Fragen NT II	241	2,52	1,34	2,00	2,00	3,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Richtige SF-Fragen VT	241	1,44	1,23	1,00	1,00	2,00
Richtige SF-Fragen NT I	241	2,53	1,44	3,00	1,00	4,00
Richtige SF-Fragen NT II	241	2,28	1,44	2,00	1,00	3,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Richtige F-Fragen VT	241	2,32	1,36	2,00	1,00	3,00
Richtige F-Fragen NT I	241	3,51	1,42	4,00	2,00	5,00
Richtige F-Fragen NT II	241	2,79	1,36	3,00	2,00	4,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige F-Fragen VT	81	2,33	1,30	2,00	1,00	3,00
		Richtige F-Fragen NT I	81	2,96	1,42	3,00	2,00	4,00
		Richtige F-Fragen NT II	81	2,54	1,24	2,00	2,00	3,50
	SF	Richtige F-Fragen VT	81	2,37	1,30	2,00	1,00	3,00
		Richtige F-Fragen NT I	81	3,54	1,34	4,00	2,50	4,50
		Richtige F-Fragen NT II	81	2,99	1,50	3,00	2,00	4,00
	F	Richtige F-Fragen VT	79	2,27	1,49	2,00	1,00	4,00
		Richtige F-Fragen NT I	79	4,03	1,30	4,00	3,00	5,00
		Richtige F-Fragen NT II	79	2,84	1,31	3,00	2,00	4,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige SF-Fragen VT	81	1,44	1,30	1,00	,00	2,00
		Richtige SF-Fragen NT I	81	2,01	1,23	2,00	1,00	3,00
		Richtige SF-Fragen NT II	81	1,88	1,32	2,00	1,00	3,00
	SF	Richtige SF-Fragen VT	81	1,40	1,18	1,00	1,00	2,00
		Richtige SF-Fragen NT I	81	3,44	1,22	3,00	3,00	4,00
		Richtige SF-Fragen NT II	81	2,79	1,46	3,00	2,00	4,00
	F	Richtige SF-Fragen VT	79	1,48	1,22	1,00	,00	2,00
		Richtige SF-Fragen NT I	79	2,11	1,41	2,00	1,00	3,00
		Richtige SF-Fragen NT II	79	2,18	1,40	2,00	1,00	3,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige S-Fragen VT	81	1,62	1,03	1,00	1,00	2,00
		Richtige S-Fragen NT I	81	2,52	1,19	2,00	2,00	3,00
		Richtige S-Fragen NT II	81	2,51	1,39	3,00	2,00	3,00
	SF	Richtige S-Fragen VT	81	1,60	1,19	1,00	1,00	2,50
		Richtige S-Fragen NT I	81	2,81	1,13	3,00	2,00	3,50
		Richtige S-Fragen NT II	81	2,72	1,38	3,00	2,00	4,00
	F	Richtige S-Fragen VT	79	1,57	1,13	1,00	1,00	2,00
		Richtige S-Fragen NT I	79	2,72	1,20	3,00	2,00	4,00
		Richtige S-Fragen NT II	79	2,33	1,23	2,00	2,00	3,00

KogMit18 Lindenhof-Kenner
(1.2.1.1.2.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	40	6,70	3,80	6,50	3,00	10,00
		Richtig NT I	40	9,13	3,53	9,00	6,00	12,00
		Richtig NT II	40	8,40	3,84	8,00	6,00	11,00
	SF	Richtig VT	38	6,11	3,13	6,00	3,75	9,00
		Richtig NT I	38	10,03	2,63	10,00	8,00	12,00
		Richtig NT II	38	9,63	2,79	9,50	8,00	11,00
	F	Richtig VT	43	5,72	2,76	5,00	4,00	7,00
		Richtig NT I	43	9,49	2,86	9,00	8,00	12,00
		Richtig NT II	43	9,00	3,55	9,00	6,00	13,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Richtige S-Fragen VT	121	1,78	1,36	1,00	1,00	3,00
Richtige S-Fragen NT I	121	3,12	1,37	3,00	2,00	4,00
Richtige S-Fragen NT II	121	3,01	1,46	3,00	2,00	4,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Richtige SF-Fragen VT	121	1,69	1,38	1,00	1,00	3,00
Richtige SF-Fragen NT I	121	2,79	1,49	3,00	2,00	4,00
Richtige SF-Fragen NT II	121	2,69	1,48	3,00	2,00	4,00

	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Richtige F-Fragen VT	121	2,69	1,38	3,00	2,00	4,00
Richtige F-Fragen NT I	121	3,63	1,35	4,00	3,00	5,00
Richtige F-Fragen NT II	121	3,30	1,36	3,00	2,00	4,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige F-Fragen VT	40	2,85	1,51	3,00	2,00	4,00
		Richtige F-Fragen NT I	40	3,30	1,45	3,50	2,00	5,00
		Richtige F-Fragen NT II	40	2,93	1,35	3,00	2,00	4,00
	SF	Richtige F-Fragen VT	38	2,61	1,35	3,00	2,00	3,00

		Richtige F-Fragen NT I	38	3,37	1,05	3,00	3,00	4,00
		Richtige F-Fragen NT II	38	3,39	1,22	3,00	3,00	4,00
	F	Richtige F-Fragen VT	43	2,63	1,31	3,00	2,00	4,00
		Richtige F-Fragen NT I	43	4,16	1,34	4,00	3,00	5,00
		Richtige F-Fragen NT II	43	3,56	1,45	3,00	2,00	5,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige SF-Fragen VT	40	1,90	1,43	1,50	1,00	3,00
		Richtige SF-Fragen NT I	40	2,45	1,65	2,00	1,00	4,00
		Richtige SF-Fragen NT II	40	2,25	1,60	2,00	1,00	3,00
	SF	Richtige SF-Fragen VT	38	1,74	1,43	2,00	,00	3,00
		Richtige SF-Fragen NT I	38	3,74	1,03	4,00	3,00	5,00
		Richtige SF-Fragen NT II	38	3,26	1,25	3,00	2,00	4,00
	F	Richtige SF-Fragen VT	43	1,47	1,30	1,00	,00	2,00
		Richtige SF-Fragen NT I	43	2,28	1,32	2,00	1,00	3,00
		Richtige SF-Fragen NT II	43	2,60	1,43	2,00	2,00	4,00

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtige S-Fragen VT	40	1,95	1,54	2,00	1,00	3,00
		Richtige S-Fragen NT I	40	3,38	1,37	3,00	2,25	4,75
		Richtige S-Fragen NT II	40	3,23	1,69	3,00	2,00	4,75
	SF	Richtige S-Fragen VT	38	1,76	1,32	1,50	1,00	3,00
		Richtige S-Fragen NT I	38	2,92	1,40	3,00	2,00	4,00
		Richtige S-Fragen NT II	38	2,97	1,30	3,00	2,00	4,00
	F	Richtige S-Fragen VT	43	1,63	1,22	1,00	1,00	2,00
		Richtige S-Fragen NT I	43	3,05	1,33	3,00	2,00	4,00
		Richtige S-Fragen NT II	43	2,84	1,38	3,00	2,00	4,00

KogO
(1.2.1.2.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Lindenhof bekannt	ja	Summenscore VT	121	6,12	4,36	6,00	3,00	8,00
		Summenscore NT I	121	11,22	5,69	11,00	7,00	15,00
		Summenscore NT II	121	9,53	5,80	9,00	5,00	13,00
	nein	Summenscore VT	241	4,78	3,25	4,00	2,00	7,00
		Summenscore NT I	241	10,30	5,75	10,00	6,00	14,00
		Summenscore NT II	241	9,24	6,22	9,00	5,00	13,00

(1.2.1.2.1.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Summenscore VT	81	4,93	3,48	5,00	2,00	7,00
		Summenscore NT I	81	11,62	5,92	12,00	8,00	15,00
		Summenscore NT II	81	9,94	6,39	9,00	5,00	15,00
	SF	Summenscore VT	81	4,77	3,37	4,00	2,00	7,00
		Summenscore NT I	81	9,80	5,37	10,00	6,00	13,00
		Summenscore NT II	81	8,68	6,09	9,00	4,00	12,50
	F	Summenscore VT	79	4,65	2,90	4,00	3,00	6,00
		Summenscore NT I	79	9,47	5,78	10,00	5,00	14,00
		Summenscore NT II	79	9,11	6,20	9,00	4,00	13,00

(1.2.1.2.2.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Summenscore VT	40	6,25	4,58	5,50	4,00	7,75
		Summenscore NT I	40	13,48	5,76	12,50	9,00	16,75
		Summenscore NT II	40	10,38	5,82	10,00	6,25	14,00
	SF	Summenscore VT	38	5,89	3,33	6,00	3,00	8,00
		Summenscore NT I	38	9,74	4,84	9,50	5,00	13,00
		Summenscore NT II	38	8,92	5,25	8,50	4,00	12,00
	F	Summenscore VT	43	6,21	5,01	5,00	3,00	9,00
		Summenscore NT I	43	10,44	5,79	9,00	6,00	15,00
		Summenscore NT II	43	9,28	6,27	9,00	4,00	13,00

KogMit18 Gesamtgruppe

(1.2.2.1.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Richtig VT	177	5,19	3,01	5,00	3,00	7,00
		Richtig NT I	177	8,54	2,99	9,00	6,00	10,50
		Richtig NT II	177	7,66	3,15	8,00	6,00	10,00
	Junge	Richtig VT	188	6,06	3,00	6,00	4,00	8,00
		Richtig NT I	188	9,44	2,95	10,00	7,00	11,00
		Richtig NT II	188	8,45	3,58	8,00	6,00	11,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	63	5,54	3,30	5,00	3,00	7,00
		Richtig NT I	63	7,73	3,30	8,00	6,00	10,00
		Richtig NT II	63	7,38	3,63	7,00	4,00	10,00
	SF	Richtig VT	53	5,09	2,81	5,00	3,00	6,00
		Richtig NT I	53	9,62	2,47	10,00	8,50	11,50
		Richtig NT II	53	8,09	2,63	8,00	7,00	10,00
	F	Richtig VT	61	4,92	2,89	5,00	2,50	7,00
		Richtig NT I	61	8,43	2,84	9,00	6,50	10,00
		Richtig NT II	61	7,57	3,05	7,00	5,00	9,50

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	60	6,12	3,19	5,50	4,00	9,00
		Richtig NT I	60	8,37	3,08	8,00	6,00	10,00
		Richtig NT II	60	7,48	3,35	7,50	6,00	10,00
	SF	Richtig VT	67	6,06	3,02	6,00	4,00	8,00
		Richtig NT I	67	10,12	2,80	10,00	9,00	12,00
		Richtig NT II	67	9,46	3,78	10,00	8,00	12,00
	F	Richtig VT	61	6,00	2,82	6,00	4,00	8,00
		Richtig NT I	61	9,74	2,71	10,00	7,00	12,00
		Richtig NT II	61	8,28	3,34	8,00	6,00	10,00

Lindenhof-Nicht-Kenner

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Richtig VT	119	4,87	2,78	5,00	3,00	7,00
		Richtig NT I	119	8,17	3,03	8,00	6,00	10,00
		Richtig NT II	119	7,03	3,04	7,00	5,00	9,00
	Junge	Richtig VT	121	5,81	2,93	5,00	4,00	8,00
		Richtig NT I	121	9,23	2,76	9,00	7,00	11,00
		Richtig NT II	121	8,12	3,45	8,00	6,00	10,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	44	4,84	2,79	4,50	3,00	6,75
		Richtig NT I	44	7,07	3,09	7,00	5,00	8,00
		Richtig NT II	44	6,70	3,41	7,00	4,00	9,75
	SF	Richtig VT	35	5,40	2,92	5,00	3,00	7,00
		Richtig NT I	35	9,77	2,66	10,00	9,00	12,00
		Richtig NT II	35	7,89	2,95	8,00	6,00	9,00
	F	Richtig VT	40	4,45	2,62	4,00	2,00	6,00
		Richtig NT I	40	7,98	2,73	8,00	6,00	10,00
		Richtig NT II	40	6,63	2,58	7,00	5,00	8,00

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	36	5,97	2,88	5,50	4,00	8,00
		Richtig NT I	36	7,89	2,52	8,00	6,00	9,75
		Richtig NT II	36	7,11	2,95	7,00	6,00	9,00
	SF	Richtig VT	46	5,35	2,84	5,00	3,00	8,00
		Richtig NT I	46	9,83	2,73	10,00	8,75	11,25
		Richtig NT II	46	8,96	4,00	9,00	6,75	12,00
	F	Richtig VT	39	6,21	3,08	6,00	4,00	8,00
		Richtig NT I	39	9,77	2,64	10,00	7,00	12,00
		Richtig NT II	39	8,08	2,97	8,00	6,00	9,00

Lindenhof -Kenner

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Richtig VT	58	5,84	3,38	5,00	3,00	8,25
		Richtig NT I	58	9,29	2,78	9,00	7,75	11,25
		Richtig NT II	58	8,97	3,00	9,00	7,00	11,00
	Junge	Richtig VT	63	6,46	3,12	6,00	4,00	9,00
		Richtig NT I	63	9,76	3,25	10,00	7,00	12,00
		Richtig NT II	63	9,03	3,84	9,00	7,00	11,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	19	7,16	3,85	7,00	5,00	11,00
		Richtig NT I	19	9,26	3,33	9,00	6,00	12,00
		Richtig NT II	19	8,95	3,72	10,00	6,00	12,00
	SF	Richtig VT	18	4,50	2,55	4,00	3,00	6,00
		Richtig NT I	18	9,33	2,09	9,00	8,00	11,00
		Richtig NT II	18	8,50	1,89	8,50	7,00	10,00
	F	Richtig VT	21	5,81	3,23	5,00	3,00	8,00
		Richtig NT I	21	9,29	2,90	9,00	8,00	12,00
		Richtig NT II	21	9,38	3,12	9,00	6,50	13,00

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Richtig VT	21	6,29	3,80	5,00	3,00	10,00
		Richtig NT I	21	9,00	3,78	8,00	5,50	12,00
		Richtig NT II	21	7,90	3,97	8,00	5,50	10,50
	SF	Richtig VT	20	7,55	2,95	7,50	5,25	10,00
		Richtig NT I	20	10,65	2,94	11,00	8,50	12,75
		Richtig NT II	20	10,65	3,12	11,00	8,25	12,75
	F	Richtig VT	22	5,64	2,30	5,50	4,00	7,00
		Richtig NT I	22	9,68	2,88	10,00	7,00	11,25
		Richtig NT II	22	8,64	3,96	8,00	6,00	12,25

KogO Gesamtgruppe
(1.2.2.2.)

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Summenscore VT	177	5,79	3,86	5,00	3,00	8,00
		Summenscore NT I	177	11,79	6,10	12,00	7,00	15,00
		Summenscore NT II	177	10,61	6,16	10,00	6,00	14,00
	Junge	Summenscore VT	188	4,69	3,52	4,00	2,00	6,00
		Summenscore NT I	188	9,53	5,13	9,00	6,00	13,00
		Summenscore NT II	188	8,11	5,72	8,00	3,00	12,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	63	6,25	3,87	6,00	3,00	9,00
		Summenscore NT I	63	13,35	6,15	13,00	9,00	16,00
		Summenscore NT II	63	11,59	5,88	11,00	8,00	17,00
	SF	Summenscore VT	53	5,08	3,41	5,00	2,00	8,00
		Summenscore NT I	53	11,06	5,27	12,00	6,00	14,00
		Summenscore NT II	53	9,83	6,19	9,00	5,00	13,50
	F	Summenscore VT	61	5,93	4,19	5,00	3,50	8,00
		Summenscore NT I	61	10,80	6,48	11,00	6,00	15,00
		Summenscore NT II	61	10,28	6,39	10,00	5,00	14,00

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	60	4,35	3,86	4,00	2,00	5,75
		Summenscore NT I	60	11,00	5,38	10,50	7,25	14,00
		Summenscore NT II	60	8,30	6,03	8,00	3,00	13,00
	SF	Summenscore VT	67	5,21	3,38	5,00	3,00	7,00
		Summenscore NT I	67	8,87	4,96	9,00	5,00	13,00
		Summenscore NT II	67	7,99	5,39	9,00	3,00	12,00
	F	Summenscore VT	61	4,46	3,31	4,00	2,00	6,00
		Summenscore NT I	61	8,82	4,83	9,00	5,00	13,00
		Summenscore NT II	61	8,07	5,85	7,00	4,00	11,00

Lindenhof-Nicht-Kenner

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Summenscore VT	119	5,26	3,44	5,00	3,00	7,00
		Summenscore NT I	119	11,34	5,96	11,00	7,00	14,00
		Summenscore NT II	119	10,31	6,08	10,00	6,00	14,00
	Junge	Summenscore VT	121	4,25	2,93	4,00	2,00	6,00
		Summenscore NT I	121	9,21	5,30	9,00	5,00	13,00
		Summenscore NT II	121	8,13	6,18	8,00	3,00	12,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	44	5,70	3,72	5,00	3,00	8,75
		Summenscore NT I	44	12,52	5,99	12,00	8,25	15,75
		Summenscore NT II	44	10,73	6,06	9,50	6,25	16,75
	SF	Summenscore VT	35	5,00	3,69	4,00	2,00	8,00
		Summenscore NT I	35	11,37	5,46	12,00	6,00	14,00
		Summenscore NT II	35	10,09	6,50	9,00	5,00	14,00
	F	Summenscore VT	40	5,00	2,89	5,00	3,00	6,00
		Summenscore NT I	40	10,00	6,19	10,50	5,25	14,00
		Summenscore NT II	40	10,05	5,85	10,00	5,25	14,00

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	36	3,78	2,68	3,50	2,00	5,00
		Summenscore NT I	36	10,28	5,57	10,00	6,00	13,75
		Summenscore NT II	36	8,78	6,67	8,00	3,00	14,75
	SF	Summenscore VT	46	4,59	3,14	4,00	2,75	7,00
		Summenscore NT I	46	8,61	5,04	8,50	4,75	12,00
		Summenscore NT II	46	7,61	5,59	9,00	2,00	11,25
	F	Summenscore VT	39	4,28	2,90	4,00	2,00	6,00
		Summenscore NT I	39	8,92	5,35	9,00	4,00	13,00
		Summenscore NT II	39	8,15	6,47	7,00	3,00	11,00

Lindenhof-Kenner

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Summenscore VT	58	6,88	4,44	6,00	4,00	9,00
		Summenscore NT I	58	12,71	6,34	12,00	7,75	17,00
		Summenscore NT II	58	11,22	6,34	11,00	7,00	15,25
	Junge	Summenscore VT	63	5,43	4,20	5,00	3,00	7,00
		Summenscore NT I	63	9,86	4,65	9,00	6,00	13,00
		Summenscore NT II	63	7,97	4,80	8,00	4,00	11,00

Nur Mädchen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	19	7,53	4,01	7,00	5,00	9,00
		Summenscore NT I	19	15,26	6,23	13,00	10,00	19,00
		Summenscore NT II	19	13,58	5,04	14,00	10,00	17,00
	SF	Summenscore VT	18	5,22	2,88	5,00	3,00	7,25
		Summenscore NT I	18	10,44	4,98	11,00	5,75	12,75
		Summenscore NT II	18	9,33	5,71	9,00	4,00	12,00
	F	Summenscore VT	21	7,71	5,59	7,00	4,00	11,50
		Summenscore NT I	21	12,33	6,89	12,00	6,50	17,50
		Summenscore NT II	21	10,71	7,44	9,00	4,00	17,50

Nur Jungen

			Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Summenscore VT	21	5,10	4,86	4,00	2,00	6,00
		Summenscore NT I	21	11,86	4,88	11,00	8,50	15,00
		Summenscore NT II	21	7,48	4,96	8,00	3,50	10,50
	SF	Summenscore VT	20	6,50	3,65	6,50	3,25	8,75
		Summenscore NT I	20	9,10	4,73	9,00	5,00	13,00
		Summenscore NT II	20	8,55	4,93	8,50	4,25	12,00
	F	Summenscore VT	22	4,77	4,00	4,00	2,00	6,00
		Summenscore NT I	22	8,64	3,85	9,00	5,75	12,00
		Summenscore NT II	22	7,91	4,69	7,00	4,00	10,50

Affektive Aussageebene
(2.1.1.)

	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Vortest						
Maus	366	3,45	1,08	4,00	3,00	4,00
Möwe	366	3,06	,89	3,00	3,00	4,00
Biber	366	3,57	,93	4,00	3,00	4,00
Storch	366	3,37	,95	3,00	3,00	4,00
Forelle	366	2,94	1,12	3,00	2,00	4,00
Schwalbe	366	3,33	,93	3,00	3,00	4,00
Goldfisch	366	3,75	1,02	4,00	3,00	5,00
Hase	366	4,63	,70	5,00	4,00	5,00
Specht	366	3,39	,82	3,00	3,00	4,00
Dachs	366	3,33	,95	3,00	3,00	4,00
Sumpfdotterschnepfe	366	2,74	1,04	3,00	2,00	3,00
Meerschweinchen	366	4,45	,85	5,00	4,00	5,00
Fischotter	366	3,33	1,04	3,00	3,00	4,00
Wellensittich	366	4,12	,96	4,00	4,00	5,00
Nachtest I						
Maus	366	4,02	,93	4,00	3,00	5,00
Möwe	366	3,40	,88	3,00	3,00	4,00
Biber	366	3,78	,95	4,00	3,00	5,00
Storch	366	3,67	,88	4,00	3,00	4,00
Forelle	366	3,19	1,07	3,00	3,00	4,00
Schwalbe	366	3,57	,93	4,00	3,00	4,00
Goldfisch	366	3,88	,98	4,00	3,00	5,00
Hase	366	4,64	,67	5,00	4,00	5,00
Specht	366	3,63	,84	4,00	3,00	4,00
Dachs	366	3,63	,98	4,00	3,00	4,00
Sumpfdotterschnepfe	366	3,16	,96	3,00	3,00	4,00
Meerschweinchen	366	4,43	,85	5,00	4,00	5,00
Fischotter	366	3,42	1,07	3,00	3,00	4,00
Wellensittich	366	4,20	,99	5,00	4,00	5,00
Nachtest II						

Möwe	366	3,32	,86	3,00	3,00	4,00
Biber	366	3,72	,92	4,00	3,00	4,00
Storch	366	3,55	,89	4,00	3,00	4,00
Forelle	366	3,20	1,07	3,00	3,00	4,00
Schwalbe	366	3,54	,92	4,00	3,00	4,00
Goldfisch	366	3,86	,93	4,00	3,00	5,00
Hase	366	4,55	,76	5,00	4,00	5,00
Specht	366	3,56	,83	4,00	3,00	4,00
Dachs	366	3,64	,97	4,00	3,00	4,00
Sumpfdotter Schnepfe	366	3,17	,98	3,00	3,00	4,00
Meerschweinchen	366	4,47	,84	5,00	4,00	5,00
Fischotter	366	3,58	1,03	4,00	3,00	4,00
Wellensittich	366	4,15	1,00	4,00	4,00	5,00

(2.1.1.1.)

Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Lindenhof gefallen	366	4,45	,74	5,00	4,00	5,00
Urwald	366	3,84	,86	4,00	3,00	4,00
Sandgrube	366	3,88	,85	4,00	3,00	4,00
Wachholderheide	366	3,78	,92	4,00	3,00	4,00
Kulturfolger	366	3,93	,96	4,00	3,00	5,00
Platzbedarf	366	3,90	,97	4,00	3,00	5,00
Archaeopteryx	366	3,93	1,07	4,00	3,00	5,00

Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Lindenhof gefallen	366	4,14	,88	4,00	4,00	5,00
Urwald	366	3,71	,85	4,00	3,00	4,00
Sandgrube	366	3,66	,97	4,00	3,00	4,00
Wachholderheide	366	3,55	1,07	4,00	3,00	4,00
Kulturfolger	366	3,76	1,01	4,00	3,00	5,00
Platzbedarf	366	3,76	,98	4,00	3,00	4,00
Archaeopteryx	366	3,90	1,05	4,00	3,00	5,00

Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Urwald	366	3,84	,86	4,00	3,00	4,00
Vögel bei Urwald	366	4,12	,82	4,00	4,00	5,00
Tiere	366	4,09	,95	4,00	4,00	5,00
Pflanzen	366	3,84	,92	4,00	3,00	4,00
Pilze	366	3,39	1,19	4,00	3,00	4,00
Landschaft	366	4,39	,84	5,00	4,00	5,00
Unordnung	366	3,52	1,44	4,00	3,00	5,00
Arbeitsblatt	366	3,48	1,27	4,00	3,00	4,00
Platzbedarf	366	3,90	,97	4,00	3,00	5,00
Vögel bei Platzbedarf	366	4,37	,79	5,00	4,00	5,00
Grafik	366	3,37	1,46	4,00	3,00	4,00
Erklärungstext	366	3,57	1,33	4,00	3,00	4,00
Arbeitsblatt	366	3,78	1,19	4,00	3,00	5,00

Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Urwald	366	3,71	,85	4,00	3,00	4,00
Vögel bei Urwald	366	3,88	1,04	4,00	3,00	5,00
Tiere	366	4,02	1,09	4,00	4,00	5,00
Pflanzen	366	3,69	1,15	4,00	3,00	4,00
Pilze	366	3,38	1,23	4,00	3,00	4,00
Landschaft	366	4,13	1,05	4,00	4,00	5,00
Unordnung	366	3,46	1,58	4,00	3,00	5,00
Arbeitsblatt	366	3,11	1,55	3,00	2,00	4,00
Platzbedarf	366	3,76	,98	4,00	3,00	4,00
Vögel bei Platzbedarf	366	4,04	1,05	4,00	4,00	5,00
Grafik	366	3,25	1,54	4,00	3,00	4,00
Erklärungstext	366	3,47	1,26	4,00	3,00	4,00
Arbeitsblatt	366	3,49	1,45	4,00	3,00	5,00

		Nachtest I	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	124	4,47	,69	5,00	4,00	5,00
		Urwald	124	3,75	,87	4,00	3,00	4,00
		Sandgrube	124	3,83	,91	4,00	3,00	4,00
		Wachholderheide	124	3,68	1,00	4,00	3,00	4,00
		Kulturfolger	124	4,05	,96	4,00	3,00	5,00
		Platzbedarf	124	3,86	,96	4,00	3,00	5,00
		Archaeopteryx	124	3,95	1,08	4,00	3,00	5,00
	SF	Lindenhof gefallen	120	4,47	,75	5,00	4,00	5,00
		Urwald	120	3,85	,85	4,00	3,00	4,00
		Sandgrube	120	3,94	,86	4,00	3,50	5,00
		Wachholderheide	120	3,79	,81	4,00	3,00	4,00
		Kulturfolger	120	3,79	,92	4,00	3,00	4,00
		Platzbedarf	120	3,81	,95	4,00	3,00	5,00
		Archaeopteryx	120	3,86	1,12	4,00	3,00	5,00
	F	Lindenhof gefallen	122	4,43	,78	5,00	4,00	5,00
		Urwald	122	3,92	,84	4,00	3,00	5,00
		Sandgrube	122	3,87	,78	4,00	3,00	4,00
		Wachholderheide	122	3,88	,94	4,00	3,00	5,00
		Kulturfolger	122	3,95	1,00	4,00	3,00	5,00
		Platzbedarf	122	4,03	1,00	4,00	3,00	5,00
		Archaeopteryx	122	3,97	1,01	4,00	3,00	5,00

		Nachtest II	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	124	4,13	,85	4,00	4,00	5,00
		Urwald	124	3,61	,86	4,00	3,00	4,00
		Sandgrube	124	3,61	,96	4,00	3,00	4,00
		Wachholderheide	124	3,59	1,10	4,00	3,00	4,00
		Kulturfolger	124	3,88	,95	4,00	3,00	5,00
		Platzbedarf	124	3,75	,93	4,00	3,00	4,00
		Archaeopteryx	124	3,91	1,05	4,00	3,00	5,00
	SF	Lindenhof gefallen	120	4,15	,92	4,00	4,00	5,00
		Urwald	120	3,73	,92	4,00	3,00	4,00
		Sandgrube	120	3,67	1,04	4,00	3,00	4,00
		Wachholderheide	120	3,59	1,04	4,00	3,00	4,00

		Kulturfolger	120	3,63	1,04	4,00	3,00	4,00
		Platzbedarf	120	3,57	1,06	4,00	3,00	4,00
		Archaeopteryx	120	3,99	1,06	4,00	3,00	5,00
	F	Lindenhof gefallen	122	4,15	,87	4,00	4,00	5,00
		Urwald	122	3,81	,73	4,00	3,00	4,00
		Sandgrube	122	3,72	,91	4,00	3,00	4,00
		Wachholderheide	122	3,46	1,05	4,00	3,00	4,00
		Kulturfolger	122	3,77	1,03	4,00	3,00	5,00
		Platzbedarf	122	3,98	,92	4,00	3,50	5,00
		Archaeopteryx	122	3,78	1,05	4,00	3,00	5,00

(2.2.1.1.)

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Lindenhof bekannt	ja	Lindenhof gefallen	121	4,36	,82	5,00	4,00	5,00
		Urwald	121	3,74	,90	4,00	3,00	4,00
		Sandgrube	121	3,83	,86	4,00	3,00	4,00
		Wachholderheide	121	3,62	,95	4,00	3,00	4,00
		Kulturfolger	121	3,91	1,01	4,00	3,00	5,00
		Platzbedarf	121	3,97	1,00	4,00	3,00	5,00
		Archaeopteryx	121	3,84	1,02	4,00	3,00	5,00
	nein	Lindenhof gefallen	241	4,50	,69	5,00	4,00	5,00
		Urwald	241	3,88	,84	4,00	3,00	4,00
		Sandgrube	241	3,89	,85	4,00	3,00	4,00
		Wachholderheide	241	3,86	,90	4,00	3,00	5,00
		Kulturfolger	241	3,94	,94	4,00	3,00	5,00
		Platzbedarf	241	3,86	,96	4,00	3,00	5,00
		Archaeopteryx	241	3,98	1,08	4,00	3,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Lindenhof bekannt	ja	Lindenhof gefallen	121	4,04	,77	4,00	4,00	5,00
		Urwald	121	3,77	,77	4,00	3,00	4,00
		Sandgrube	121	3,66	,98	4,00	3,00	4,00
		Wachholderheide	121	3,65	1,04	4,00	3,00	4,00

		Kulturfolger	121	3,82	,98	4,00	3,00	5,00
		Platzbedarf	121	3,75	,97	4,00	3,00	5,00
		Archaeopteryx	121	3,90	1,04	4,00	3,00	5,00
	nein	Lindenhof gefallen	241	4,19	,93	4,00	4,00	5,00
		Urwald	241	3,68	,89	4,00	3,00	4,00
		Sandgrube	241	3,67	,97	4,00	3,00	4,00
		Wachholderheide	241	3,49	1,08	4,00	3,00	4,00
		Kulturfolger	241	3,73	1,02	4,00	3,00	4,00
		Platzbedarf	241	3,76	,99	4,00	3,00	4,00
		Archaeopteryx	241	3,89	1,06	4,00	3,00	5,00

Lindenhof-Nicht-Kenner

		Nachtest I	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	81	4,45	,68	5,00	4,00	5,00
		Urwald	81	3,70	,85	4,00	3,00	4,00
		Sandgrube	81	3,85	,92	4,00	3,00	4,75
		Wachholderheide	81	3,74	1,02	4,00	3,00	5,00
		Kulturfolger	81	4,00	,91	4,00	3,00	5,00
		Platzbedarf	81	3,80	,89	4,00	3,00	4,00
		Archaeopteryx	81	3,98	1,13	4,00	3,00	5,00
	SF	Lindenhof gefallen	81	4,51	,77	5,00	4,00	5,00
		Urwald	81	3,94	,83	4,00	3,00	5,00
		Sandgrube	81	3,93	,87	4,00	3,00	5,00
		Wachholderheide	81	3,91	,79	4,00	3,00	5,00
		Kulturfolger	81	3,81	,90	4,00	3,00	4,00
		Platzbedarf	81	3,73	,97	4,00	3,00	4,00
		Archaeopteryx	81	3,98	1,15	4,00	3,00	5,00
	F	Lindenhof gefallen	79	4,54	,62	5,00	4,00	5,00
		Urwald	79	4,00	,81	4,00	3,00	5,00
		Sandgrube	79	3,90	,77	4,00	3,00	4,00
		Wachholderheide	79	3,93	,88	4,00	3,00	5,00
		Kulturfolger	79	4,00	,99	4,00	3,00	5,00
		Platzbedarf	79	4,05	1,00	4,00	3,00	5,00
		Archaeopteryx	79	3,99	,96	4,00	3,00	5,00

		Nachtest II	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	81	4,14	,92	4,00	4,00	5,00
		Urwald	81	3,49	,90	4,00	3,00	4,00
		Sandgrube	81	3,53	,96	4,00	3,00	4,00
		Wachholderheide	81	3,47	1,11	4,00	3,00	4,00
		Kulturfolger	81	3,78	,94	4,00	3,00	4,75
		Platzbedarf	81	3,75	,96	4,00	3,00	4,00
		Archaeopteryx	81	3,85	1,05	4,00	3,00	5,00
	SF	Lindenhof gefallen	81	4,20	,97	4,00	4,00	5,00
		Urwald	81	3,77	1,00	4,00	3,00	4,00
		Sandgrube	81	3,71	1,08	4,00	3,00	4,00
		Wachholderheide	81	3,62	1,09	4,00	3,00	4,00
		Kulturfolger	81	3,62	1,10	4,00	3,00	4,00
		Platzbedarf	81	3,53	1,06	4,00	3,00	4,00
		Archaeopteryx	81	4,03	1,08	4,00	3,00	5,00
	F	Lindenhof gefallen	79	4,24	,92	4,00	4,00	5,00
		Urwald	79	3,79	,71	4,00	3,00	4,00
		Sandgrube	79	3,76	,85	4,00	3,00	4,00
		Wachholderheide	79	3,38	1,04	3,00	3,00	4,00
		Kulturfolger	79	3,81	1,03	4,00	3,00	5,00
		Platzbedarf	79	4,01	,91	4,00	3,50	5,00
		Archaeopteryx	79	3,78	1,06	4,00	3,00	5,00

Lindenhof-Kenner

		Nachtest I	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	40	4,46	,73	5,00	4,00	5,00
		Urwald	40	3,79	,92	4,00	3,00	4,00
		Sandgrube	40	3,75	,93	4,00	3,00	4,00
		Wachholderheide	40	3,51	,97	3,00	3,00	4,00
		Kulturfolger	40	4,08	1,06	4,00	3,00	5,00
		Platzbedarf	40	3,92	1,09	4,00	3,00	5,00
		Archaeopteryx	40	3,90	1,01	4,00	3,00	5,00
	SF	Lindenhof gefallen	38	4,39	,72	5,00	4,00	5,00
		Urwald	38	3,66	,91	4,00	3,00	4,00
		Sandgrube	38	3,94	,86	4,00	4,00	4,75

		Wachholderheide	38	3,54	,82	4,00	3,00	4,00
		Kulturfolger	38	3,80	,93	4,00	3,00	4,00
		Platzbedarf	38	4,00	,88	4,00	3,00	5,00
		Archaeopteryx	38	3,68	,94	4,00	3,00	4,00
	F	Lindenhof gefallen	43	4,23	,97	4,00	4,00	5,00
		Urwald	43	3,76	,89	4,00	3,00	4,00
		Sandgrube	43	3,80	,81	4,00	3,00	4,00
		Wachholderheide	43	3,79	1,02	4,00	3,00	5,00
		Kulturfolger	43	3,86	1,03	4,00	4,00	4,25
		Platzbedarf	43	3,98	1,02	4,00	3,00	5,00
		Archaeopteryx	43	3,93	1,10	4,00	3,00	5,00

		Nachtest II	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	40	4,10	,72	4,00	4,00	5,00
		Urwald	40	3,82	,79	4,00	3,00	4,00
		Sandgrube	40	3,74	,97	4,00	3,00	4,00
		Wachholderheide	40	3,81	1,09	4,00	3,00	5,00
		Kulturfolger	40	4,08	,97	4,00	3,75	5,00
		Platzbedarf	40	3,72	,92	4,00	3,00	4,00
		Archaeopteryx	40	4,00	1,07	4,00	3,00	5,00
	SF	Lindenhof gefallen	38	4,03	,83	4,00	4,00	5,00
		Urwald	38	3,63	,75	4,00	3,00	4,00
		Sandgrube	38	3,61	,97	4,00	3,00	4,00
		Wachholderheide	38	3,51	,95	4,00	3,00	4,00
		Kulturfolger	38	3,68	,91	4,00	3,00	4,00
		Platzbedarf	38	3,61	1,05	3,50	3,00	5,00
		Archaeopteryx	38	3,95	1,03	4,00	3,50	5,00
	F	Lindenhof gefallen	43	4,00	,77	4,00	4,00	4,50
		Urwald	43	3,85	,77	4,00	3,00	4,00
		Sandgrube	43	3,63	1,02	4,00	3,00	4,00
		Wachholderheide	43	3,62	1,06	4,00	3,00	4,00
		Kulturfolger	43	3,69	1,03	4,00	3,00	4,00
		Platzbedarf	43	3,93	,94	4,00	3,25	5,00
		Archaeopteryx	43	3,78	1,05	4,00	3,00	4,75

Gesamtgruppe
(2.2.2.)

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Geschlecht	Mädchen	Lindenhof gefallen	177	4,57	,68	5,00	4,00	5,00
		Urwald	177	3,84	,79	4,00	3,00	4,00
		Sandgrube	177	3,91	,78	4,00	3,00	4,00
		Wachholderheide	177	3,86	,87	4,00	3,00	5,00
		Kulturfolger	177	4,11	,91	4,00	4,00	5,00
		Platzbedarf	177	3,99	,95	4,00	3,00	5,00
		Archaeopteryx	177	3,85	1,00	4,00	3,00	5,00
	Junge	Lindenhof gefallen	188	4,34	,78	5,00	4,00	5,00
		Urwald	188	3,82	,91	4,00	3,00	4,00
		Sandgrube	188	3,84	,92	4,00	3,00	4,00
		Wachholderheide	188	3,71	,97	4,00	3,00	4,00
		Kulturfolger	188	3,76	,98	4,00	3,00	5,00
		Platzbedarf	188	3,81	,99	4,00	3,00	5,00
		Archaeopteryx	188	3,99	1,13	4,00	3,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Geschlecht	Mädchen	Lindenhof gefallen	177	4,36	,71	4,00	4,00	5,00
		Urwald	177	3,81	,73	4,00	3,00	4,00
		Sandgrube	177	3,74	,88	4,00	3,00	4,00
		Wachholderheide	177	3,73	,96	4,00	3,00	4,00
		Kulturfolger	177	4,01	,86	4,00	3,00	5,00
		Platzbedarf	177	3,91	,91	4,00	3,00	5,00
		Archaeopteryx	177	3,80	1,03	4,00	3,00	5,00
	Junge	Lindenhof gefallen	188	3,92	,98	4,00	3,00	5,00
		Urwald	188	3,62	,93	4,00	3,00	4,00
		Sandgrube	188	3,59	1,05	4,00	3,00	4,00
		Wachholderheide	188	3,36	1,13	4,00	3,00	4,00
		Kulturfolger	188	3,52	1,08	4,00	3,00	4,00
		Platzbedarf	188	3,62	1,03	4,00	3,00	4,00
		Archaeopteryx	188	3,98	1,07	4,00	3,00	5,00

Nur Mädchen

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	63	4,62	,49	5,00	4,00	5,00
		Urwald	63	3,79	,75	4,00	3,00	4,00
		Sandgrube	63	3,89	,79	4,00	3,00	4,00
		Wachholderheide	63	3,89	,90	4,00	3,00	5,00
		Kulturfolger	63	4,29	,85	4,00	4,00	5,00
		Platzbedarf	63	4,06	,88	4,00	3,00	5,00
		Archaeopteryx	63	3,92	,99	4,00	3,00	5,00
	SF	Lindenhof gefallen	53	4,57	,73	5,00	4,00	5,00
		Urwald	53	3,87	,81	4,00	3,00	4,00
		Sandgrube	53	4,00	,82	4,00	4,00	5,00
		Wachholderheide	53	3,91	,81	4,00	3,00	4,50
		Kulturfolger	53	4,02	,84	4,00	3,00	5,00
		Platzbedarf	53	3,89	,91	4,00	3,00	5,00
		Archaeopteryx	53	4,02	,90	4,00	3,00	5,00
	F	Lindenhof gefallen	61	4,51	,80	5,00	4,00	5,00
		Urwald	61	3,88	,83	4,00	3,00	4,00
		Sandgrube	61	3,85	,76	4,00	3,00	4,00
		Wachholderheide	61	3,78	,88	4,00	3,00	4,00
		Kulturfolger	61	4,00	1,02	4,00	4,00	5,00
		Platzbedarf	61	4,00	1,05	4,00	3,00	5,00
		Archaeopteryx	61	3,63	1,07	4,00	3,00	4,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	63	4,35	,74	4,00	4,00	5,00
		Urwald	63	3,71	,81	4,00	3,00	4,00
		Sandgrube	63	3,65	,90	4,00	3,00	4,00
		Wachholderheide	63	3,73	,90	4,00	3,00	4,00
		Kulturfolger	63	4,11	,84	4,00	4,00	5,00
		Platzbedarf	63	3,89	,91	4,00	3,00	5,00
		Archaeopteryx	63	3,88	1,06	4,00	3,00	5,00
	SF	Lindenhof gefallen	53	4,36	,69	4,00	4,00	5,00
		Urwald	53	3,86	,69	4,00	4,00	4,00
		Sandgrube	53	3,92	,84	4,00	3,00	5,00

		Wachholderheide	53	3,92	,90	4,00	3,00	5,00
		Kulturfolger	53	3,80	,92	4,00	3,00	4,00
		Platzbedarf	53	3,80	,96	4,00	3,00	5,00
		Archaeopteryx	53	3,94	,93	4,00	3,00	5,00
	F	Lindenhof gefallen	61	4,36	,71	4,00	4,00	5,00
		Urwald	61	3,88	,66	4,00	3,50	4,00
		Sandgrube	61	3,69	,87	4,00	3,00	4,00
		Wachholderheide	61	3,56	1,06	3,50	3,00	4,00
		Kulturfolger	61	4,09	,81	4,00	3,50	5,00
		Platzbedarf	61	4,04	,85	4,00	4,00	5,00
		Archaeopteryx	61	3,58	1,05	4,00	3,00	4,00

Nur Jungen

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	60	4,28	,83	4,00	4,00	5,00
		Urwald	60	3,68	,97	4,00	3,00	4,00
		Sandgrube	60	3,75	1,03	4,00	3,00	4,00
		Wachholderheide	60	3,46	1,07	3,00	3,00	4,00
		Kulturfolger	60	3,78	1,01	4,00	3,00	5,00
		Platzbedarf	60	3,62	,99	4,00	3,00	4,00
		Archaeopteryx	60	3,97	1,18	4,00	3,00	5,00
	SF	Lindenhof gefallen	67	4,39	,77	5,00	4,00	5,00
		Urwald	67	3,84	,89	4,00	3,00	4,00
		Sandgrube	67	3,89	,90	4,00	3,00	5,00
		Wachholderheide	67	3,69	,80	4,00	3,00	4,00
		Kulturfolger	67	3,60	,94	4,00	3,00	4,00
		Platzbedarf	67	3,75	,98	4,00	3,00	5,00
		Archaeopteryx	67	3,73	1,26	4,00	3,00	5,00
	F	Lindenhof gefallen	61	4,35	,76	4,00	4,00	5,00
		Urwald	61	3,95	,86	4,00	3,00	5,00
		Sandgrube	61	3,88	,81	4,00	3,00	4,00
		Wachholderheide	61	3,98	,98	4,00	3,00	5,00
		Kulturfolger	61	3,90	1,00	4,00	3,00	5,00
		Platzbedarf	61	4,05	,96	4,00	3,00	5,00
		Archaeopteryx	61	4,30	,83	4,50	4,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	60	3,88	,90	4,00	3,75	4,00
		Urwald	60	3,47	,90	4,00	3,00	4,00
		Sandgrube	60	3,56	1,04	4,00	3,00	4,00
		Wachholderheide	60	3,40	1,27	4,00	2,00	4,00
		Kulturfolger	60	3,63	1,01	4,00	3,00	4,00
		Platzbedarf	60	3,58	,93	4,00	3,00	4,00
		Archaeopteryx	60	3,93	1,04	4,00	3,00	5,00
	SF	Lindenhof gefallen	67	3,98	1,05	4,00	3,25	5,00
		Urwald	67	3,63	1,06	4,00	3,00	4,00
		Sandgrube	67	3,48	1,13	4,00	3,00	4,00
		Wachholderheide	67	3,32	1,08	3,00	3,00	4,00
		Kulturfolger	67	3,50	1,11	4,00	3,00	4,00
		Platzbedarf	67	3,38	1,10	4,00	3,00	4,00
		Archaeopteryx	67	4,03	1,15	4,00	3,00	5,00
	F	Lindenhof gefallen	61	3,90	,98	4,00	3,00	5,00
		Urwald	61	3,75	,79	4,00	3,00	4,00
		Sandgrube	61	3,75	,96	4,00	3,00	4,00
		Wachholderheide	61	3,36	1,04	4,00	3,00	4,00
		Kulturfolger	61	3,43	1,13	3,50	3,00	4,00
		Platzbedarf	61	3,93	,98	4,00	3,00	5,00
		Archaeopteryx	61	3,98	1,03	4,00	3,00	5,00

Lindenhof-Nicht-Kenner

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Lindenhof gefallen	119	4,65	,58	5,00	4,00	5,00
		Urwald	119	3,92	,79	4,00	3,00	4,25
		Sandgrube	119	3,96	,79	4,00	3,00	5,00
		Wachholderheide	119	4,02	,82	4,00	3,00	5,00
		Kulturfolger	119	4,12	,85	4,00	4,00	5,00
		Platzbedarf	119	3,98	,96	4,00	3,00	5,00
		Archaeopteryx	119	3,98	,96	4,00	3,00	5,00
	Junge	Lindenhof gefallen	121	4,36	,76	5,00	4,00	5,00

		Sandgrube	121	3,82	,90	4,00	3,00	4,00
		Wachholderheide	121	3,70	,96	4,00	3,00	4,00
		Kulturfolger	121	3,75	,99	4,00	3,00	5,00
		Platzbedarf	121	3,72	,95	4,00	3,00	4,00
		Archaeopteryx	121	3,97	1,18	4,00	3,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Lindenhof gefallen	119	4,45	,73	5,00	4,00	5,00
		Urwald	119	3,84	,75	4,00	3,00	4,00
		Sandgrube	119	3,77	,84	4,00	3,00	4,00
		Wachholderheide	119	3,74	,94	4,00	3,00	4,00
		Kulturfolger	119	3,99	,86	4,00	3,00	5,00
		Platzbedarf	119	3,96	,85	4,00	3,00	5,00
		Archaeopteryx	119	3,90	,98	4,00	3,00	5,00
	Junge	Lindenhof gefallen	121	3,90	1,04	4,00	3,00	5,00
		Urwald	121	3,52	,98	4,00	3,00	4,00
		Sandgrube	121	3,56	1,07	4,00	3,00	4,00
		Wachholderheide	121	3,23	1,15	3,00	2,00	4,00
		Kulturfolger	121	3,47	1,11	4,00	3,00	4,00
		Platzbedarf	121	3,55	1,07	4,00	3,00	4,00
		Archaeopteryx	121	3,86	1,14	4,00	3,00	5,00

Nur Mädchen

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	44	4,60	,50	5,00	4,00	5,00
		Urwald	44	3,77	,77	4,00	3,00	4,00
		Sandgrube	44	3,86	,82	4,00	3,00	4,75
		Wachholderheide	44	4,00	,91	4,00	3,00	5,00
		Kulturfolger	44	4,14	,88	4,00	4,00	5,00
		Platzbedarf	44	3,98	,90	4,00	3,00	5,00
		Archaeopteryx	44	3,89	1,04	4,00	3,00	5,00
	SF	Lindenhof gefallen	35	4,64	,74	5,00	4,50	5,00
		Urwald	35	4,00	,80	4,00	4,00	5,00
		Sandgrube	35	4,06	,85	4,00	4,00	5,00
		Wachholderheide	35	4,17	,66	4,00	4,00	5,00
		Kulturfolger	35	4,12	,74	4,00	4,00	5,00

		Platzbedarf	35	3,83	,98	4,00	3,00	5,00
		Archaeopteryx	35	4,32	,77	4,50	4,00	5,00
	F	Lindenhof gefallen	40	4,72	,51	5,00	4,25	5,00
		Urwald	40	4,03	,78	4,00	3,00	5,00
		Sandgrube	40	3,98	,70	4,00	3,25	4,00
		Wachholderheide	40	3,90	,84	4,00	3,00	4,75
		Kulturfolger	40	4,10	,91	4,00	4,00	5,00
		Platzbedarf	40	4,13	,99	4,00	3,00	5,00
		Archaeopteryx	40	3,79	,98	4,00	3,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	44	4,39	,81	5,00	4,00	5,00
		Urwald	44	3,59	,84	4,00	3,00	4,00
		Sandgrube	44	3,57	,93	4,00	3,00	4,00
		Wachholderheide	44	3,60	,94	4,00	3,00	4,00
		Kulturfolger	44	3,95	,89	4,00	3,00	5,00
		Platzbedarf	44	3,84	,95	4,00	3,00	5,00
		Archaeopteryx	44	3,90	1,01	4,00	3,00	5,00
	SF	Lindenhof gefallen	35	4,52	,62	5,00	4,00	5,00
		Urwald	35	4,06	,61	4,00	4,00	4,00
		Sandgrube	35	4,03	,85	4,00	4,00	5,00
		Wachholderheide	35	4,16	,77	4,00	4,00	5,00
		Kulturfolger	35	3,88	,93	4,00	3,00	5,00
		Platzbedarf	35	3,91	,88	4,00	3,50	4,50
		Archaeopteryx	35	4,16	,88	4,00	4,00	5,00
	F	Lindenhof gefallen	40	4,47	,72	5,00	4,00	5,00
		Urwald	40	3,95	,66	4,00	4,00	4,00
		Sandgrube	40	3,78	,70	4,00	3,00	4,00
		Wachholderheide	40	3,54	,99	3,00	3,00	4,00
		Kulturfolger	40	4,14	,75	4,00	4,00	5,00
		Platzbedarf	40	4,16	,69	4,00	4,00	5,00
		Archaeopteryx	40	3,68	1,00	4,00	3,00	4,50

Nur Jungen

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	36	4,26	,83	4,00	4,00	5,00
		Urwald	36	3,57	,92	4,00	3,00	4,00
		Sandgrube	36	3,80	1,02	4,00	3,00	4,00
		Wachholderheide	36	3,42	1,08	3,00	3,00	4,00
		Kulturfolger	36	3,80	,93	4,00	3,00	5,00
		Platzbedarf	36	3,54	,82	4,00	3,00	4,00
		Archaeopteryx	36	4,06	1,24	5,00	3,00	5,00
	SF	Lindenhof gefallen	46	4,42	,79	5,00	4,00	5,00
		Urwald	46	3,89	,85	4,00	3,00	5,00
		Sandgrube	46	3,83	,88	4,00	3,00	4,00
		Wachholderheide	46	3,70	,82	4,00	3,00	4,00
		Kulturfolger	46	3,57	,94	4,00	3,00	4,00
		Platzbedarf	46	3,66	,96	4,00	3,00	4,00
		Archaeopteryx	46	3,72	1,31	4,00	3,00	5,00
	F	Lindenhof gefallen	39	4,37	,67	4,00	4,00	5,00
		Urwald	39	3,97	,85	4,00	3,00	5,00
		Sandgrube	39	3,82	,83	4,00	3,00	4,00
		Wachholderheide	39	3,97	,94	4,00	3,00	5,00
		Kulturfolger	39	3,90	1,07	4,00	3,00	5,00
		Platzbedarf	39	3,97	1,01	4,00	3,00	5,00
		Archaeopteryx	39	4,18	,91	4,00	4,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	36	3,80	,96	4,00	3,00	4,00
		Urwald	36	3,33	,93	3,00	3,00	4,00
		Sandgrube	36	3,47	1,03	4,00	3,00	4,00
		Wachholderheide	36	3,26	1,26	3,50	2,00	4,00
		Kulturfolger	36	3,54	,98	4,00	3,00	4,00
		Platzbedarf	36	3,61	,96	4,00	3,00	4,00
		Archaeopteryx	36	3,74	1,09	4,00	3,00	5,00
	SF	Lindenhof gefallen	46	3,95	1,11	4,00	4,00	5,00
		Urwald	46	3,57	1,17	4,00	3,00	4,00
		Sandgrube	46	3,48	1,17	4,00	3,00	4,00

		Wachholderheide	46	3,21	1,14	3,00	2,75	4,00
		Kulturfolger	46	3,43	1,19	4,00	3,00	4,00
		Platzbedarf	46	3,25	1,10	4,00	3,00	4,00
		Archaeopteryx	46	3,93	1,19	4,00	3,00	5,00
	F	Lindenhof gefallen	39	3,94	1,06	4,00	3,00	5,00
		Urwald	39	3,64	,72	4,00	3,00	4,00
		Sandgrube	39	3,75	1,00	4,00	3,00	4,75
		Wachholderheide	39	3,20	1,08	3,00	3,00	4,00
		Kulturfolger	39	3,46	1,17	3,00	3,00	4,00
		Platzbedarf	39	3,86	1,07	4,00	3,00	5,00
		Archaeopteryx	39	3,89	1,13	4,00	3,00	5,00

Lindenhof-Kenner

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Lindenhof gefallen	58	4,41	,82	5,00	4,00	5,00
		Urwald	58	3,68	,79	4,00	3,00	4,00
		Sandgrube	58	3,81	,77	4,00	3,00	4,00
		Wachholderheide	58	3,53	,87	3,00	3,00	4,00
		Kulturfolger	58	4,09	1,04	4,00	4,00	5,00
		Platzbedarf	58	4,00	,94	4,00	3,00	5,00
		Archaeopteryx	58	3,58	1,02	4,00	3,00	4,00
	Junge	Lindenhof gefallen	63	4,30	,83	4,00	4,00	5,00
		Urwald	63	3,80	1,00	4,00	3,00	5,00
		Sandgrube	63	3,85	,95	4,00	3,00	5,00
		Wachholderheide	63	3,71	1,02	4,00	3,00	4,00
		Kulturfolger	63	3,75	,96	4,00	3,00	4,00
		Platzbedarf	63	3,93	1,06	4,00	3,00	5,00
		Archaeopteryx	63	4,08	,97	4,00	3,00	5,00

		Nachtest II	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Geschlecht	Mädchen	Lindenhof gefallen	58	4,16	,65	4,00	4,00	5,00
		Urwald	58	3,75	,69	4,00	3,00	4,00
		Sandgrube	58	3,69	,94	4,00	3,00	4,00
		Wachholderheide	58	3,72	1,01	4,00	3,00	4,50
		Kulturfolger	58	4,05	,87	4,00	3,50	5,00

		Platzbedarf	58	3,80	1,00	4,00	3,00	5,00
		Archaeopteryx	58	3,59	1,09	4,00	3,00	4,00
	Junge	Lindenhof gefallen	63	3,93	,86	4,00	3,00	5,00
		Urwald	63	3,78	,85	4,00	3,00	4,00
		Sandgrube	63	3,63	1,02	4,00	3,00	4,00
		Wachholderheide	63	3,58	1,07	4,00	3,00	4,00
		Kulturfolger	63	3,58	1,03	4,00	3,00	4,00
		Platzbedarf	63	3,70	,95	4,00	3,00	4,00
		Archaeopteryx	63	4,21	,90	4,00	4,00	5,00

Nur Mädchen

		Nachtest I	Anzahl	Mittelwert	Standardabweichung	Median	25. Perzentil	75. Perzentil
Treatment	S	Lindenhof gefallen	19	4,68	,48	5,00	4,00	5,00
		Urwald	19	3,83	,71	4,00	3,00	4,00
		Sandgrube	19	3,95	,71	4,00	3,00	4,00
		Wachholderheide	19	3,63	,83	4,00	3,00	4,00
		Kulturfolger	19	4,63	,68	5,00	4,00	5,00
		Platzbedarf	19	4,26	,81	4,00	4,00	5,00
		Archaeopteryx	19	4,00	,88	4,00	3,00	5,00
	SF	Lindenhof gefallen	18	4,44	,70	5,00	4,00	5,00
		Urwald	18	3,61	,78	4,00	3,00	4,00
		Sandgrube	18	3,89	,76	4,00	3,75	4,00
		Wachholderheide	18	3,39	,85	3,00	3,00	4,00
		Kulturfolger	18	3,83	,99	4,00	3,00	5,00
		Platzbedarf	18	4,00	,77	4,00	3,00	5,00
		Archaeopteryx	18	3,44	,86	3,50	3,00	4,00
	F	Lindenhof gefallen	21	4,14	1,06	4,00	3,50	5,00
		Urwald	21	3,60	,88	4,00	3,00	4,00
		Sandgrube	21	3,60	,82	4,00	3,00	4,00
		Wachholderheide	21	3,55	,94	3,50	3,00	4,00
		Kulturfolger	21	3,80	1,20	4,00	3,25	5,00
		Platzbedarf	21	3,76	1,14	4,00	3,00	5,00
		Archaeopteryx	21	3,30	1,17	3,00	3,00	4,00

		Nachtest II	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	19	4,26	,56	4,00	4,00	5,00
		Urwald	19	4,00	,67	4,00	4,00	4,00
		Sandgrube	19	3,84	,83	4,00	3,00	4,00
		Wachholderheide	19	4,06	,73	4,00	3,75	5,00
		Kulturfolger	19	4,47	,61	5,00	4,00	5,00
		Platzbedarf	19	4,00	,82	4,00	3,00	5,00
		Archaeopteryx	19	3,83	1,20	4,00	2,75	5,00
	SF	Lindenhof gefallen	18	4,06	,75	4,00	4,00	4,50
		Urwald	18	3,50	,71	4,00	3,00	4,00
		Sandgrube	18	3,72	,83	4,00	3,00	4,00
		Wachholderheide	18	3,50	,99	3,50	3,00	4,00
		Kulturfolger	18	3,67	,91	4,00	3,00	4,00
		Platzbedarf	18	3,61	1,09	3,50	3,00	5,00
		Archaeopteryx	18	3,56	,92	4,00	3,00	4,00
	F	Lindenhof gefallen	21	4,14	,65	4,00	4,00	5,00
		Urwald	21	3,75	,64	4,00	3,00	4,00
		Sandgrube	21	3,52	1,12	3,00	3,00	4,50
		Wachholderheide	21	3,59	1,23	4,00	3,00	5,00
		Kulturfolger	21	4,00	,92	4,00	3,00	5,00
		Platzbedarf	21	3,79	1,08	4,00	3,00	5,00
		Archaeopteryx	21	3,40	1,14	4,00	2,25	4,00

Nur Jungen

		Nachtest I	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	21	4,22	,88	4,00	4,00	5,00
		Urwald	21	3,76	1,09	4,00	3,00	5,00
		Sandgrube	21	3,57	1,08	4,00	3,00	4,50
		Wachholderheide	21	3,40	1,10	3,00	3,00	4,00
		Kulturfolger	21	3,55	1,10	4,00	3,00	4,00
		Platzbedarf	21	3,60	1,23	4,00	3,00	5,00
		Archaeopteryx	21	3,81	1,12	4,00	3,00	5,00
	SF	Lindenhof gefallen	20	4,35	,75	4,50	4,00	5,00

		Sandgrube	20	4,00	,97	4,00	3,75	5,00
		Wachholderheide	20	3,71	,77	4,00	3,00	4,00
		Kulturfolger	20	3,76	,90	4,00	3,00	4,00
		Platzbedarf	20	4,00	1,00	4,00	3,00	5,00
		Archaeopteryx	20	3,89	,99	4,00	3,00	5,00
	F	Lindenhof gefallen	22	4,32	,89	4,50	4,00	5,00
		Urwald	22	3,90	,89	4,00	3,00	5,00
		Sandgrube	22	4,00	,77	4,00	3,00	5,00
		Wachholderheide	22	4,00	1,07	4,00	3,75	5,00
		Kulturfolger	22	3,91	,87	4,00	4,00	4,00
		Platzbedarf	22	4,19	,87	4,00	4,00	5,00
		Archaeopteryx	22	4,52	,60	5,00	4,00	5,00

		Nachtest II	An- zahl	Mittelwert	Standardabweichung	Median	25. Per- zentil	75. Per- zentil
Treatment	S	Lindenhof gefallen	21	3,95	,83	4,00	3,25	4,75
		Urwald	21	3,65	,88	4,00	3,00	4,00
		Sandgrube	21	3,65	1,09	4,00	3,00	4,75
		Wachholderheide	21	3,56	1,34	4,00	2,00	5,00
		Kulturfolger	21	3,68	1,11	4,00	3,00	5,00
		Platzbedarf	21	3,45	,94	3,00	3,00	4,00
		Archaeopteryx	21	4,17	,92	4,00	3,75	5,00
	SF	Lindenhof gefallen	20	4,00	,92	4,00	3,00	5,00
		Urwald	20	3,75	,79	4,00	3,25	4,00
		Sandgrube	20	3,50	1,10	4,00	3,00	4,00
		Wachholderheide	20	3,53	,94	4,00	3,00	4,00
		Kulturfolger	20	3,68	,95	4,00	3,00	4,00
		Platzbedarf	20	3,60	1,05	3,50	3,00	4,75
		Archaeopteryx	20	4,32	1,00	5,00	4,00	5,00
	F	Lindenhof gefallen	22	3,85	,88	4,00	3,25	4,00
		Urwald	22	3,95	,89	4,00	3,00	5,00
		Sandgrube	22	3,75	,91	4,00	3,00	4,00
		Wachholderheide	22	3,65	,93	4,00	3,00	4,00
		Kulturfolger	22	3,37	1,07	4,00	3,00	4,00
		Platzbedarf	22	4,05	,80	4,00	4,00	5,00
		Archaeopteryx	22	4,15	,81	4,00	4,00	5,00

Dank

Ich möchte mich herzlich bei denen bedanken, die zur Fertigstellung der Arbeit beigetragen haben, v. a. allen Mitgliedern des Lehrstuhls Didaktik der Biologie an der Universität von Bayreuth.

Zuerst danke ich Prof. Dr. **Siegfried Klautke** für die Überlassung dieses spannenden Themas, seine stete Anteilnahme und Ansprechbarkeit und die vielen anregenden Diskussionen.

Aus Tradition und Verdienst kommt sofort danach Frau **Sabine Hübner**. Sie war mir eine große Hilfe bei der statistischen Auswertung und „Vertraute in stürmischer Zeit“.

Ich danke meinen beiden Kollegen Dr. **Reinhard Tutschek** und OStR **Franz-Josef Scharfenberg**. Auf sehr unterschiedliche Art und Weise hatten beide großen Anteil am Gelingen der Arbeit.

Weiter möchte ich mich bei den beteiligten Wissenschaftlichen Hilfskräften, den Schülern und Lehrern, die am Museumsunterricht teilnahmen, den Mitarbeitern des Lindenhofs und den Teilnehmern der 3. bis 5. Frühjahrsschule bedanken.

Last but not least danke ich Prof. Dr. **Franz Bogner** für seine Geduld (und seine Ungeduld).

Erklärung

Hiermit erkläre ich, dass ich die Arbeit selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ferner erkläre ich, dass ich anderweitig mit oder ohne Erfolg nicht versucht habe, diese Dissertation einzureichen. Ich habe keine gleichartige Doktorprüfung an einer anderen Hochschule endgültig nicht bestanden.

Bayreuth, den 07.06.2004

(Matthias Wilde)